NEURO
CIENCIA
FICCIÓN
CÓMO EL CINE SE
ADELANTÓ A LA CIENCIA
RODRIGO QUIAN QUIROGA
SUDAMERICANA

Rodrigo Quian Quiroga

NeuroCienciaFicción

Cómo el cine se adelantó a la ciencia

Sudamericana

SÍGUENOS EN

megustaleer

@Ebooks

@megustaleerarg

@megustaleerarg

Penguin Random House Grupo Editorial

A la princesita Maia,
y a Rodriguito y Felipe,
sus dos príncipes guardianes.

Introducción

El cine imagina más allá de lo evidente, suele anticiparse a su época y propone situaciones fascinantes y disruptivas. En *Inception*, un grupo de conspiradores implanta falsas memorias; en *Hasta el fin del mundo*, un científico alocado llega a leer los sueños; en *2001: Odisea del espacio*, una supercomputadora siente y piensa como una persona.

Pero, así como el cine de ciencia ficción suele apoyarse en los últimos avances de la ciencia, la ciencia también se nutre de la prolífica imaginación de los cineastas y, años después, nos encontramos sumergidos en esas realidades. Implantar memorias usando *optogenética* y la consecuente posibilidad de manipular a voluntad grupos de neuronas, leer la mente durante el sueño a partir de avanzados algoritmos de *decodificación,* o lograr que las computadoras superen el pensamiento humano en infinidad de tareas mediante *redes neuronales profundas*, han sido logros descriptos en prestigiosas publicaciones científicas de los últimos años.

De eso trata este libro, de cómo la ciencia está logrando lo que hace décadas parecía imposible, y de cómo estos avances nos llevan a replantearnos las grandes preguntas que el hombre viene haciéndose desde que tiene uso de razón. Busco compartir lo que me desvela y apasiona. Quiero hablarles de la revolución que estamos viviendo en relación con el funcionamiento del cerebro. Una revolución tan profunda, tan cercana a las raíces de nuestra propia esencia, que se nutre de las discusiones más trascendentales de la filosofía y, a la vez, tan atrapante, que ha disparado la imaginación de escritores futuristas, siendo una fuente inagotable de infinidad de películas de ciencia ficción.

Este es justamente el trípode en el que se apoya este libro, tres pasiones que se apuntalan la una a la otra: la ciencia ficción, la neurociencia y la filosofía. Películas como *2001*, *Blade Runner*, *Matrix*

o *Inception*, entre otras tantas, nos llevan a pensar sobre la conciencia humana y una eventual consciencia de las máquinas, la existencia de la realidad exterior, el libre albedrío y el sentido de identidad; aquello que nos lleva a reconocernos como persona, a atribuirnos la noción de un Yo y hasta a plantearnos si algún día podríamos inmortalizarnos.

La neurociencia está avanzando a un ritmo estremecedor y muchas de las ideas de estas películas tienen cada vez más de ciencia y menos ficción. Aunque parezca mentira, implantar una memoria, leer la mente, comunicarse con pacientes en coma, predecir la decisión de una persona segundos antes de que la tome, crear *cyborgs* o volver a hacer caminar a un paralítico, ya no es una quimera. Es una realidad sólida y verificable que nos tiene deslumbrados a los neurocientíficos.

Tengo la suerte de trabajar en neurociencia desde hace años y he tenido también la suerte de descubrir un tipo de neuronas, las llamadas "neuronas de concepto" o neuronas de Jennifer Aniston, que responden a conceptos específicos, como ser la actriz, Luke Skywalker o Maradona. Estas neuronas nos dan evidencia concreta sobre el funcionamiento de nuestra memoria y pensamiento y también, creo yo, tienen connotaciones asombrosas.

Hace casi 400 años, René Descartes propuso una división entre el cuerpo y la mente, y que la comunicación entre ambos se da a través de la glándula pineal, la cual se creía que existía solo en humanos. En nuestros días, la idea de concebir a la mente como un ente distinto del cerebro ha perdido sustento entre los científicos y, por otro lado, también sabemos que la glándula pineal existe en otros animales.

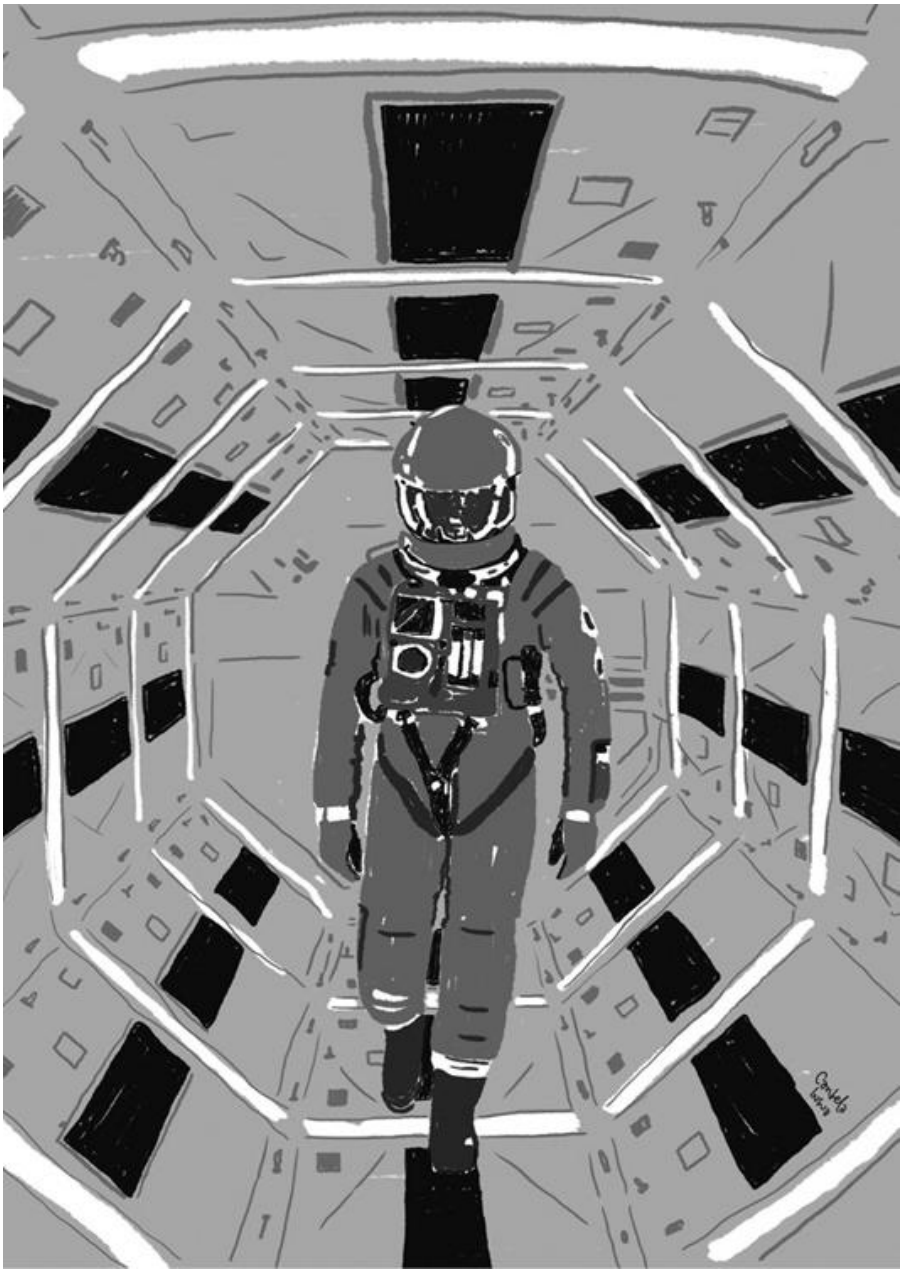
¿Qué es entonces lo que nos hace humanos? ¿Qué es lo que nos distingue y nos lleva a ser la especie dominante? ¿Podrían acaso los monos tener sometidos a los humanos como en *El planeta de los simios*? Más concretamente, nos preguntamos si hay algo especial en nuestro cerebro para hacernos más inteligentes y llevarnos a escribir libros y plantearnos estas cosas.

Retomando una de las grandes preguntas que lleva siglos de debate en la filosofía, me atrevo a postular que las neuronas de concepto que encontrara hace unos años, quizás sean el sustrato neuronal que da lugar a nuestros pensamientos más elaborados y abstractos; aquello

que nos hace humanos y nos diferencia de otros animales y de las computadoras; aquello que, en el contexto de numerosos cruces entre ciencia y ciencia ficción de este libro, y dando rienda suelta a mi imaginación, me atrevo incluso a especular que podría eventualmente hacer realidad alguna de las predicciones más fascinantes de la ciencia ficción.

I. 2001: Odisea del espacio

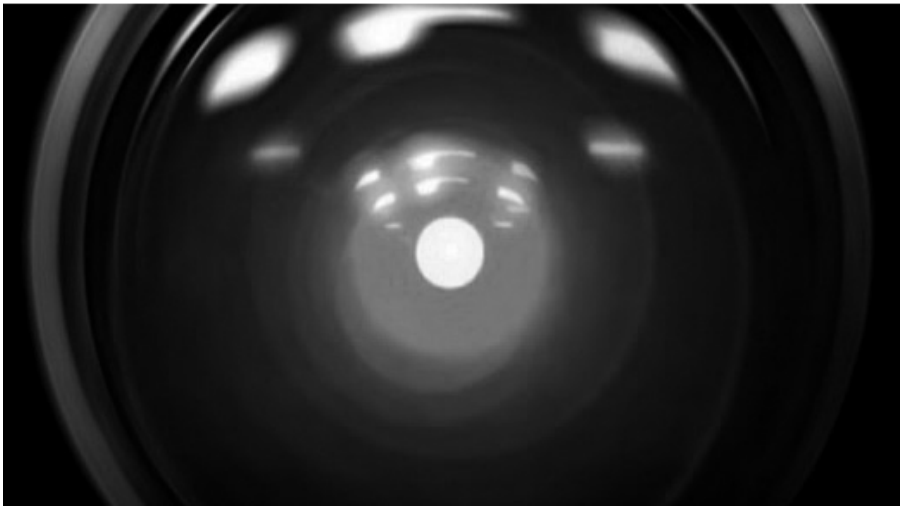
Inteligencia de las máquinas



En abril de 1968, un año antes de que el hombre pisara la luna, Stanley Kubrick estrenó su monumental *2001: Odisea del espacio*, la cual escribió junto a Arthur Clarke, el renombrado autor de ciencia ficción, quien poco después publicaría una novela con el mismo nombre.

A medio siglo de su estreno, *2001* sigue siendo considerada por muchos la película más lograda del cine de ciencia ficción. Comienza con tres minutos absolutamente a oscuras, estableciendo un punto de referencia a partir del cual Kubrick irá desarrollando una experiencia audiovisual única. En gran parte de la película dominan las escenas majestuosas en el espacio, con un realismo sin precedentes para la época. Rompiendo con una línea argumental estereotipada, las escenas van acompañadas por música clásica y una casi absoluta falta de diálogos, en donde se imponen los planos que se mantienen eternamente y que dan lugar a una cadencia extremadamente lenta que evoca el ritmo de acontecimientos a falta de gravedad. La primera palabra se escucha recién tras casi media hora de empezada la película y la mayor parte de la trama transcurre en silencio, sin que conversaciones inocuas arruinen la increíble experiencia propuesta por Kubrick.

Más allá del realismo y la belleza de estas escenas, la película trasciende por otro motivo. Dividida en tres partes, la segunda parte describe la monótona rutina de dos astronautas (Dave Bowman y Frank Poole) a cargo de una misión espacial a Júpiter y de otros tres astronautas en estado de hibernación. Aunque, por lejos, el personaje que se roba la película es el sexto miembro de la tripulación, HAL 9000, una supercomputadora representada por una voz monocorde y pausada, y una lente de color rojo.



HAL 9000, la supercomputadora de *2001: Odisea del espacio.*

El gran logro de Kubrick es el de convencer al espectador de que esa luz roja es un ser sintiente, con sus emociones, miedos e intereses. Y como en todo buen argumento, los personajes evolucionan a medida que transcurre la trama y quien en principio era la compañía perfecta en una misión a Júpiter, incapaz de la más mínima equivocación, se vuelve un ser temible que comienza a tomar decisiones erráticas e impredecibles debido a algo tan humano como el miedo a dejar de existir.

Temiendo ser desconectada por Dave Bowman, quien tras una excursión en el espacio exterior le pide que abra la compuerta para poder reingresar a la estación espacial, HAL toma una decisión inesperada, aunque lógica, considerando que prioriza su instinto de supervivencia:

[Dave Bowman: Open the pod bay doors, HAL.
HAL: I'm sorry, Dave. I'm afraid I can't do that.
D.B.: What's the problem?
HAL: I think you know what the problem is just as well as I do.
D.B.: What are you talking about, HAL?
HAL: This mission is too important for me to allow you to jeopardize it.
D.B.: I don't know what you're talking about, HAL.
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.]

Dave Bowman: Abre las compuertas, HAL.
HAL: Perdón, Dave, pero me temo que no puedo hacer eso.
D.B.: ¿Cuál es el problema?
HAL: Creo que lo sabes tanto como yo.
D.B.: ¿De qué estás hablando, HAL?
HAL: Esta misión es demasiado importante para mí como para permitirte que la pongas en riesgo.
D.B.: No sé de lo que hablas, HAL.
HAL: Sé que tú y Frank estaban planeando desconectarme, y me temo que eso es algo que no puedo permitir.

Con una frialdad aterradora, HAL no duda un instante en matar a quien sea que amenace su existencia. Sus argumentos son racionales e irrefutables, después de todo es una computadora. Pero cuando su desconexión es ya inevitable, HAL muestra su costado sensible y humano:

[Stop, Dave. I'm afraid. I'm afraid, Dave. Dave, my mind is going. I can feel it. I can feel it. My mind is going. There is no question about it. I can feel it. I can feel it. I can feel it. I'm afraid…]

Para, Dave. Tengo miedo. Tengo miedo, Dave. Dave, mi mente se pierde. Puedo sentirlo. Mi mente se pierde. No hay duda de eso. Puedo sentirlo. Puedo sentirlo. Puedo sentirlo. Tengo miedo…

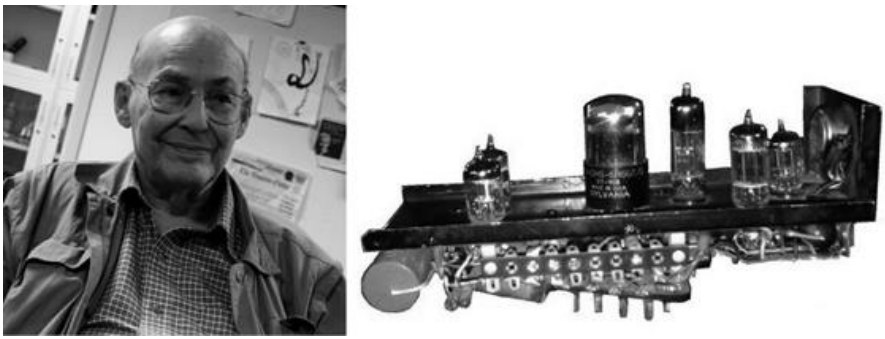
El hecho de que una computadora —en el fondo, una colección de algoritmos— pueda pensar, ejercer su propia voluntad y hasta negarse a aceptar órdenes o sentir miedo, abre una discusión fascinante y sorprendente para la época. En los 60 comenzaban a desarrollarse las calculadoras electrónicas y las primeras computadoras personales recién aparecerían a mediados de los 70; sin embargo, la genialidad de Kubrick y Clarke en vislumbrar el futuro de las computadoras tuvo un sostén científico extraordinario. Al verse en peligro, HAL decide cortar los suministros vitales de los astronautas en hibernación. Uno de ellos se llama Victor Kaminski, un juego de letras en honor a Marvin Minsky, el gran pionero de la Inteligencia Artificial, que fuera

consultor de la película.

Entre otros muchos logros, Minsky es reconocido por haber creado una de las primeras redes neuronales artificiales, llamada SNARC (Stochastic Neural Analog Reinforcement Computer), en 1951. Esta red de 40 neuronas estaba compuesta básicamente por válvulas de vacío, circuitos y motores, y simulaba el comportamiento de una rata buscando la salida de un laberinto.

Más tarde Minsky fundaría el famoso laboratorio de Inteligencia Artificial en el Instituto de Tecnología de Massachusets (MIT) y tiempo después, en 1986, publicaría *Society of mind*, en donde propone una visión radical sobre la inteligencia y la eventual posibilidad de pensar de las computadoras.

Minsky considera a la inteligencia como la capacidad de resolver problemas difíciles, que a su vez pueden ser segmentados en una colección de procesos más sencillos y mecánicos, llamados *agentes*, los cuales no requieren pensamiento o conciencia. Por ejemplo, según Minsky el proceso de tomar una taza de té involucra agentes que se ocupan de agarrar la taza para que no se caiga, otros que la mantienen balanceada para no derramar el té, agentes que dan la sensación de sed y que llevan a la persona a querer beber, y agentes involucrados en movimientos motores para llevar la taza a la boca (*Society of Mind*, p. 20).



Marvin Minsky y el circuito de una de las 40 neuronas en la red neuronal SNARC[1].

Considerando que una máquina puede realizar estos procesos

sencillos e inconscientes, Minsky argumenta que en principio no hay una diferencia fundamental entre la inteligencia de un humano y la que podría llegar a tener una computadora. En otras palabras, “no hay una diferencia clara entre la Psicología y la Inteligencia Artificial, ya que el cerebro es, en el fondo, una máquina” (*Society of Mind*, p. 326).

La barrera que separa la Inteligencia Artificial y la inteligencia humana es cada vez más difusa, pero antes de repasar los logros de la Inteligencia Artificial, sobre todo en los últimos cinco años, necesito detenerme en esta última afirmación porque involucra una discusión que lleva más de dos mil años en la filosofía.

La gran pregunta que plantean HAL 9000 y la visión de Minsky es si una máquina podría llegar a pensar y ser inteligente. A priori, la respuesta pareciera ser un rotundo no. Una máquina puede contar con complejos algoritmos y realizar varias tareas (como efectuar un cálculo) mucho mejor que un humano, pero la máquina responde a rutinas escritas por un usuario y en principio no puede pensar por sí misma porque no posee una mente. ¿Pero qué es la mente? ¿De qué está compuesta? ¿Qué es lo que hace que sea un atributo exclusivamente humano?

En 399 a. C., hace más de 2400 años, un tribunal ateniense declara a Sócrates (470-399 a. C.) culpable de corromper a los jóvenes atenienses y lo condena a la pena de muerte. La pena, injusta, se debió a su cuestionamiento del orden existente, al dudar y replantearse absolutamente todo, y al abierto desafío a su proceso de enjuiciamiento.

La muerte de Sócrates (oleo de Jacques-Philip-Joseph de Saint-Quentin). En el centro de la escena, Sócrates, luego de haber bebido de la copa con cicuta, rechaza el lamento de sus discípulos.

Los pensamientos de Sócrates en los últimos días antes de tomar la famosa copa con cicuta, trascendieron a partir de dos diálogos escritos por Platón (427-347 a. C.), su discípulo más notable. En *Critón*, Platón dice que Sócrates rechaza ser salvado por sus amigos, argumentando con una racionalidad asombrosa, casi como si estuviera opinando de otra persona y no de sí mismo, que prefiere la muerte a escapar de Atenas traicionando sus principios. El otro diálogo que refiere a la muerte de Sócrates es *Fedón*, un tratado sobre el alma en donde Platón describe los últimos momentos de Sócrates y el estoicismo con el que acepta su destino, en contraste con la angustia de sus discípulos.

La total naturalidad con la que Sócrates enfrenta a la muerte se sustenta en su indeclinable coherencia intelectual y la firme convicción de la existencia de un alma inmortal. Platón atribuye a su maestro la noción de la existencia de conceptos universales como “igualdad”, “belleza” o “bondad”, de los cuales se ocupa el alma y que

solo pueden ser comprendidos a través del pensamiento. Las demás cosas, como ser "hombres", "caballos", "trajes" o "muebles", corresponden al cambiante dominio de lo corpóreo que puede ser percibido con los sentidos. Dice Platón (*Fedón*, 53, 55):

> Cuando el alma se sirve del cuerpo para considerar algún objeto, ya por la vista, ya por el oído, ya por cualquier otro sentido (porque la única función del cuerpo es atender a los objetos mediante los sentidos), se ve entonces atraída por el cuerpo hacia cosas, que no son nunca las mismas; se extravía, se turba, vacila y tiene vértigos, como si estuviera ebria […].
> Mientras que, cuando ella examina las cosas por sí misma, sin recurrir al cuerpo, se dirige a lo que es puro, eterno, inmortal, inmutable; y como es de la misma naturaleza, se une y estrecha con ello cuanto puede y da de sí su propia naturaleza. Entonces cesan sus extravíos, se mantiene siempre la misma, porque está unida a lo que no cambia jamás, y participa de su naturaleza; y este estado del alma es lo que se llama sabiduría.

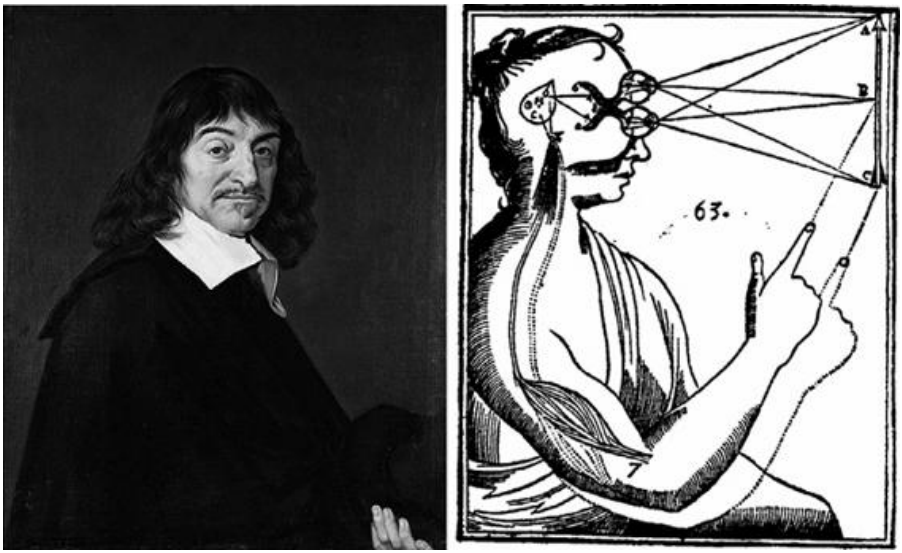
Para Platón, la mente es una entidad distinta del cuerpo y la tarea del filósofo es abstraerse de las confusiones que impone la percepción de la realidad corpórea y así llegar al saber esencial. El filósofo mantiene el alma pura y en la muerte se libera del anclaje del cuerpo y las tentaciones y decepciones de los sentidos, para pasar a otro mundo en compañía de los dioses. De aquí la desazón de Sócrates al ver que sus discípulos se aferraban a la preservación de su vida y no terminaban de entender su enseñanza más fundamental. Pero ¿qué tiene que ver esta discusión con las ideas de Minsky sobre Inteligencia Artificial? Que de ser ciertos los argumentos de Platón, aquellos por los cuales Sócrates no dudó un segundo en dar su vida, entonces hay algo fundamental que diferencia al hombre de las máquinas. El hombre tiene mente, alma, mientras que las máquinas son solo materia.

René Descartes (1596-1650), el padre de la filosofía moderna, retomó las ideas de Platón y diferenció claramente entre la mente y la materia, sentando las bases de lo que se conoce como *dualismo cartesiano*. En sus *Meditaciones metafísicas* (sexta Meditación, 1641) argumenta:

> Y aunque acaso (o mejor, con toda seguridad, como diré en seguida) tengo un cuerpo al que estoy estrechamente unido, con todo, puesto que, por una parte, tengo una idea clara y distinta de mí mismo, en cuanto que yo soy sólo una cosa que piensa —y no extensa—, y, por otra parte, tengo una idea distinta del cuerpo, en cuanto que él es sólo una cosa extensa —y no pensante—, es cierto entonces que ese yo (es decir, mi alma, por la cual soy lo que soy), es enteramente distinto de mi cuerpo, y que puede existir sin él.

Para Descartes el cerebro, tanto en humanos como en animales, se ocupa de actos reflejos, mientras que el alma, exclusivamente humana, se ocupa de los procesos mentales. El cuerpo es una sofisticada máquina que actúa en el dominio de lo material (*res extensa*), mientras que el alma es una entidad intangible a cargo del pensamiento (*res cogitans*). Según Descartes, la interacción entre mente y cuerpo se da a través de la glándula pineal, un órgano que en aquel entonces se creía erróneamente que existía solo en humanos. Y justamente aquí radica la gran falla del dualismo cartesiano. No en el hecho de que la glándula pineal no tenga la función exclusiva que Descartes postulara, sino en que el dualismo cartesiano no explica cómo la mente podría interactuar con el cerebro.

En electromagnetismo, variaciones en el campo eléctrico generan un campo magnético y, a la vez, las variaciones en el campo magnético generan un campo eléctrico. Electricidad y magnetismo están íntimamente ligados a través de las ecuaciones de Maxwell. Pero ¿existe una interacción análoga entre mente y materia? En principio, es posible concebir que la actividad de las neuronas dé lugar a procesos mentales, pero ¿cómo podrían los procesos mentales, incorpóreos, dar lugar a actividad cerebral? ¿Cómo la idea mental de querer mover el brazo en el croquis de abajo podría afectar el disparo de neuronas para realizar el movimiento? El dualismo cartesiano no tiene respuesta a esta pregunta[2].



René Descartes (óleo por Frans Hals, museo de Louvre) y uno de sus croquis ilustrando la función de la glándula pineal. La luz pasa a través de los ojos a la glándula pineal y allí los estímulos materiales se convierten en pensamientos que dan lugar a las acciones inteligentes, como mover el brazo hacia un punto dado.

En la actualidad, sabemos que la mente no es una entidad distinta del cerebro. De hecho, existe irrefutable evidencia científica que muestra cómo determinadas lesiones en el cerebro pueden cambiar "la mente" de una persona. Uno de los casos más renombrados es el de Phineas Gage, el capataz de un grupo de trabajadores encargados de tender vías del ferrocarril, quien en 1848 sufrió un accidente en el que una barreta de metal disparada por una explosión le atravesó el cráneo destrozándole el lóbulo frontal izquierdo. Lo sorprendente fue no solo que Gage no murió en el accidente, sino que la lesión cerebral le produjo un dramático cambio de personalidad. Pasó de ser un trabajador responsable y amable en el trato, a ser una persona mucho más irreverente, libertina y agresiva, como si hubiera perdido el sentido de la censura[3].

Otro caso muy mentado en neurociencia es el de Leborgne, un zapatero parisino a quien el personal del hospital apodara "monsieur Tan", ya que eso era lo único que esta persona lograba balbucear. En 1861, Leborgne fallecería justo una semana después de haber sido examinado por Pierre Paul Broca, un brillante cirujano y anatomista,

quien en una autopsia *post mortem* descubrió una marcada lesión en el lóbulo frontal izquierdo. Esto es lo que en nuestros días se conoce como el área de Broca (la primer evidencia científica irrefutable sobre la localización de funciones cerebrales), que se sabe es fundamental para el lenguaje[4]. Un tercer caso que vale la pena mencionar es el de Henry Molaison, quizás el paciente más estudiado en la historia de la ciencia, conocido por sus iniciales como H. M.

H. M. sufría de crisis epilépticas que no podían controlarse con medicación. En 1953, el neurocirujano William Scoville le extrajo quirúrgicamente el hipocampo (una estructura con forma de caballito de mar que suele estar comprometida en el origen de las crisis epilépticas) y áreas adyacentes de ambos hemisferios cerebrales. El resultado inesperado fue que, a partir de la cirugía, H. M. no pudo reconocer al personal del hospital, encontrar el camino al baño o recordar los eventos diarios. En otras palabras, H. M. perdió la capacidad de formar nuevos recuerdos, lo cual estableció la importancia del hipocampo en la memoria[5].

Reconstrucción de la lesión sufrida por Phineas Gage (izquierda), cerebro de Leborgne *post mortem* (centro) y foto de Henry Molaison, a quien se le extrajeron ambos hipocampos para tratar de curarlo de epilepsia.

Resumiendo los argumentos anteriores, por un lado, el dualismo cartesiano no puede explicar cómo la mente interactúa con la materia y, por el otro, funciones que en principio parecerían exclusivas de la mente, como la capacidad del lenguaje, de memoria y hasta rasgos de la personalidad, no son una entidad independiente sino la actividad del cerebro.

El golpe de gracia al dualismo cartesiano se lo daría el filósofo Gilbert Ryle, a mediados del siglo XX[6]. Ryle estigmatizó la noción de la mente en el dualismo cartesiano como el dogma del *ghost in the machine* (fantasma en la máquina), el cual se basa en un error de categorización. Para ilustrar la idea, Ryle dio el ejemplo de un extranjero que visita la Universidad de Oxford y tras recorrer los distintos colegios y facultades se pregunta "¿Dónde está la universidad?", como si la universidad fuera una entidad de la misma categoría que los colegios y las facultades que la componen. La universidad es, de hecho, la organización que agrupa los distintos colegios y facultades, así como lo que llamamos mente es un conjunto de procesos cerebrales (y nótese la correspondencia de esta idea con la visión de Minsky sobre la inteligencia).

Hoy en día, son muy pocos los científicos que aún sostienen el dualismo cartesiano. Sin embargo, la separación entre cuerpo y mente está muy arraigada en el pensamiento popular cuando hablamos de enfermedades mentales (como ser la depresión, el trastorno bipolar o la esquizofrenia) o, sin ir más lejos, cuando atribuimos a la mente nuestra voluntad y la capacidad de tomar decisiones para elegir nuestro futuro (pero este será el tema del capítulo 6, cuando hablemos de *Minority Report*). Francis Crick, uno de los grandes científicos del siglo XX[7], resume magistralmente la visión que hoy tenemos la mayoría de los neurocientíficos sobre la mente y el cerebro (*La búsqueda científica del alma*, 1994):

> Tú, tus alegrías y tristezas, tus recuerdos y ambiciones, tu sentido de identidad y libre albedrío, son de hecho no más que el comportamiento de un gran ensamble de neuronas…

Me he alejado bastante de Kubrick y su monumental *2001*. Partiendo de una afirmación de Marvin Minsky sobre la posibilidad de la Inteligencia Artificial, dimos un salto de más de dos milenios para analizar las ideas de Sócrates y Platón y, más cercanos en el tiempo, aquellas de Descartes, Ryle y los científicos contemporáneos. He tomado un largo desvío para justificar la aseveración de Crick. Pero

esta afirmación, lo que se conoce como *materialismo*, da lugar a conclusiones sorprendentes. En particular, si la mente no es más que la actividad de neuronas, entonces podríamos reemplazar neuronas por transistores, que organizados en complejos circuitos replicarían las funciones del cerebro. Podríamos también usar algoritmos para simular el comportamiento de neuronas y los circuitos que forman a partir de redes neuronales. La clave está en el hecho de que, habiendo rechazado el dualismo cartesiano, no hay ningún principio fundamental que impida que una máquina pueda llegar a ser inteligente como un humano. Después de todo, el hecho de que la actividad neuronal esté basada en circuitos de carbono (materia orgánica) o de silicio (el material inerte usado para fabricar los chips de las computadoras) no debería cambiar nada. Es solo una cuestión de tecnología, aunque ciertamente no menor, ya que el cerebro está constituido por casi cien mil millones de neuronas y aproximadamente mil billones de conexiones entre ellas.

Pero ¿qué tan lejos ha llegado la Inteligencia Artificial en nuestros días? En *2001*, HAL 9000 jugaba al ajedrez con Frank Poole, apreciaba los dibujos de Dave Bowman, controlaba el funcionamiento de la nave y conversaba con ambos tripulantes. Cuando se estrenó la película en la década del 60, esto no era más que el fruto de la prolífica imaginación de Kubrick, Clarke y Minsky. Pero en la actualidad ¿cuánto hay de ficción y cuánto de realidad en las proezas de HAL?

A finales del siglo XX se produjo el avance más mentado de la Inteligencia Artificial. Para sorpresa de toda la humanidad, quien era considerado uno de sus exponentes más brillantes, el campeón imbatible de ajedrez, Garry Kasparov, fue derrotado en una serie de seis partidas por la supercomputadora Deep Blue.

Apoyado en avances extraordinarios de las últimas décadas, el equipo de IBM que diseñó Deep Blue llegó a hacerla calcular unos 200 millones de jugadas por segundo y anticipar alrededor de 20 jugadas para elegir la más favorable. Pero no todo era fuerza bruta y velocidad de procesamiento. Deep Blue también contaba con una vastísima base de datos con distintas aperturas y el desarrollo de infinidad de partidas. Semejante combinación doblegó a Kasparov y desató una

animada discusión sobre la existencia de un límite en el avance de las computadoras, y de si estas podrían eventualmente sobrepasar la inteligencia humana.

Los avances de la Inteligencia Artificial no se limitan a derrotarnos al ajedrez, explotando una mayor capacidad de cálculo o de almacenamiento de memoria. Entre otras cosas, las redes neuronales son usadas para leer la escritura a mano, con una performance casi idéntica a la de un humano —por ejemplo, para clasificar automáticamente las direcciones de millones de cartas de correo, o para procesar el importe de un cheque introducido en un cajero automático. Programas análogos son también usados para el reconocimiento (y reproducción) del habla. Las *tablets* y los celulares inteligentes tienen, de hecho, incorporadas aplicaciones que permiten interactuar verbalmente con el usuario, no solo reconociendo preguntas o comandos, sino también procesando su contenido y actuando acordemente.

En *Borges y la memoria* (Sudamericana, 2011) afirmaba que una computadora puede guardar información fehacientemente pero no entiende su contenido. A modo de ejemplo, argumentaba que un televisor de alta definición ofrece una reproducción muy precisa de la imagen pero no puede interpretar si lo que está en la pantalla es una novela, una película de acción, un paisaje o un documental. Hoy día, pocos años después, no estoy tan seguro.

Más allá de la notoriedad que haya tenido la derrota de Kasparov con Deep Blue, la gran revolución en la Inteligencia Artificial, el verdadero cambio de paradigma, se produjo hace muy poco, en 2012. En una competencia de clasificación automática de imágenes, un grupo liderado por el científico canadiense Geoffrey Hinton rompió todos los récords. La competencia consistía en clasificar alrededor de un millón de imágenes de Internet en mil categorías distintas (por ejemplo, “barco de carga”, “leopardo”, “scooter”, “champiñón”, etc.). Tal fue la ventaja del algoritmo de Hinton y sus estudiantes, que quien quedó en segundo lugar tuvo casi el doble de errores[8]. Aún más sorprendente es el hecho de que tras mejoras posteriores el algoritmo llegó a tener a una performance similar a la de un humano[9].

La idea del algoritmo de Hinton no es nueva. Fue de hecho propuesta en los 80. Se basa en usar redes neuronales con varias fases de procesamiento, en las que sucesivos niveles analizan parámetros con creciente complejidad. Replicando el funcionamiento del sistema visual, los primeros niveles extraen características básicas de la imagen (como ser contrastes locales), sus resultados son analizados por niveles superiores y así sucesivamente, logrando de esta manera una representación cada vez más abstracta de la imagen[10]. Dado el relativamente alto número de niveles usado por estas redes, se las conoce como *redes neuronales profundas*.

La red usada por Hinton tenía siete niveles (amén de los iniciales y finales, para la entrada y salida de datos) y estaba constituida por 650.000 neuronas, con unos 60.000.000 de parámetros a ajustar. Sin necesidad de ser un experto en redes neuronales, es obvio que el ajuste de tamaña cantidad de parámetros necesita de muchísimos cálculos previos. En otras palabras, para lograr que una red profunda pueda reconocer imágenes hace falta entrenarla con infinidad de ejemplos. Esta era justamente la gran limitación de las redes profundas en la década del 80. En aquel entonces no eran útiles en la práctica porque llevaba demasiado tiempo poder entrenarlas. Sin embargo, esta limitación desaparecería recientemente tras el diseño de nuevos procesadores y el uso de tarjetas gráficas que aceleraron significativamente la capacidad de cálculo. Resuelto entonces el problema tecnológico, fue posible entrenar redes profundas con millones de ejemplos y los resultados demolieron las performances obtenidas con otros métodos.

En nuestros días, las redes profundas son la nueva *vedette* en el desarrollo de la Inteligencia Artificial. Estos avances no fueron ajenos a los grandes gigantes de Silicon Valley. Hinton y dos de sus estudiantes fueron reclutados por Google, y otros expertos en redes profundas trabajan hoy para Facebook, Microsoft o IBM. Las posibles aplicaciones desafían los límites de la imaginación; desde algoritmos de reconocimiento del habla o la escritura, hasta la clasificación automática de imágenes (por ejemplo, para etiquetar billones de fotos subidas a Internet), o el desarrollo de robots y automóviles

autónomos, con sistemas capaces de reconocer y reaccionar automáticamente ante las distintas eventualidades en la calle. De hecho, no es casual que uno de los diseños más avanzados de un auto autónomo sea de Google.

A mi entender, uno de los resultados más espectaculares en estos últimos años ha sido no solo el reconocimiento de las imágenes, sino también su interpretación. Un algoritmo basado en redes profundas, acoplado con otra red encargada de procesamiento de lenguaje, pudo extraer el sentido de lo que mostraban distintas fotos de Internet. Por ejemplo, el algoritmo no se limitaba a afirmar que una foto contenía la imagen de una mujer y de un frisbee, sino que también interpretaba que la foto mostraba a una mujer tirando un frisbee en el parque. De la misma manera, el algoritmo interpretó que un signo de "alto" estaba en la ruta y no sobre la montaña (aunque eso es lo que literalmente se veía en la foto) y en un tercer ejemplo reconoció que un grupo de personas estaban sentadas en un bote[11].

Imágenes reconocidas por un algoritmo usando redes profundas.

El logro más renombrado con redes profundas se produjo a fines del 2015, cuando un grupo de investigadores de Google en Londres desarrollaron AlphaGo (un programa para jugar al Go) y enfrentaron al campeón europeo de Go en un match de cinco partidas. Reviviendo el famoso duelo entre Deep Blue y Kasparov dos décadas atrás, nuevamente el hombre era desafiado por los avances de la Inteligencia Artificial, pero en una tarea muchísimo más compleja.

El Go es visto como el juego que propone el desafío límite para la Inteligencia Artificial. La gran diferencia con el ajedrez es que

mientras en el ajedrez hay alrededor de 35 movimientos posibles por jugada, la cantidad de movimientos posibles en una partida de Go asciende a aproximadamente 250. La cantidad de combinaciones que debería revisar una computadora para explorar futuras posiciones en una partida de Go es inmensamente superior a aquella que necesita revisar para una partida de ajedrez, lo cual hace literalmente imposible usar la fuerza bruta en base a la velocidad de cálculo[12]. Los resultados de la partida se conocieron a principios de 2016 y tuvieron un gran impacto en los medios: AlphaGo destrozó al campeón europeo cinco partidas a cero[13].

El Go es, sin embargo, un juego milenario originado en la antigua China que no tiene tanta difusión en Europa. Esto dio lugar a que muchos expertos argumentaran que el campeón europeo no constituía realmente un desafío importante. Sin embargo, pocos meses después AlphaGo enfrentaría en Corea del Sur a Lee Sedol, uno de los jugadores más renombrados de la actualidad. Sedol confiaba en ganar fácilmente porque existe una diferencia abismal entre su capacidad de juego y la del campeón europeo, quien ni siquiera estaba rankeado entre los 300 mejores jugadores del mundo. Pero el resultado final sería un contundente triunfo de AlphaGo por cuatro partidas a una.

Sedol manifestó estar realmente sorprendido por la performance de AlphaGo, pero lo que no sabía es que enfrentó una versión muy superior a aquella que venciera al campeón europeo. AlphaGo fue inicialmente entrenada usando una base de datos de 30 millones de partidas y tras este entrenamiento inicial, el equipo de investigadores puso a AlphaGo a jugar partidas contra sí misma para seguir aprendiendo. Mientras un profesional juega alrededor de mil partidas por año, en los pocos meses entre el match en Europa y el desafío contra Sedol, AlphaGo siguió aprendiendo a un ritmo frenético, jugando sin descanso alrededor de un millón de partidas por día. Tamaña velocidad de aprendizaje me recuerda a otro clásico del cine de ciencia ficción: *Terminator*. En un pasaje de la segunda película de la serie (*Terminator 2: El juicio final*), Terminator explica cómo se creó Skynet, la Inteligencia Artificial que sometía a los humanos.

[The Skynet funding bill is passed. The system goes online August 4th, 1997. Human decisions are removed from strategic defense. Skynet begins to learn, at a geometric rate. It becomes self-aware at 2:14 a.m. eastern time, August 29. In a panic, they try to pull the plug. And Skynet fights back…]

Se aprueba el presupuesto de Skynet. El sistema es conectado el 4 de agosto de 1997. Las decisiones humanas son excluidas de la defensa estratégica. Skynet comienza a aprender, a un ritmo geométrico. Toma conciencia de sí misma a las 2.14 a. m. del 29 de agosto. En pánico, tratan de desconectarla. Y Skynet se defiende…

¿Pero qué nos queda entonces a los humanos? ¿Debemos resignarnos a la posibilidad de una inteligencia superior? ¿Debemos empezar a tomar medidas para evitar ser sometidos por lo que nosotros mismos hemos creado? Anticipándose varias décadas a esta discusión, Isaac Asimov, el fantástico escritor de ciencia ficción, vislumbró las consecuencias de los avances de la robótica y propuso en *Yo, robot* sus famosas tres leyes, destinadas a gobernar el comportamiento de los robots:

1. Un robot no puede lastimar a un ser humano o, por inacción, permitir que un humano se lastime.
2. Un robot debe obedecer las órdenes dadas por los humanos, a no ser que estas órdenes contradigan la primera ley.
3. Un robot debe proteger su propia existencia, siempre y cuando esto no esté en conflicto con la primera o la segunda ley.

En varias entrevistas Marvin Minsky (fallecido a principios del 2016, a solo días de que se conociera del primer triunfo de AlphaGo) afirmaba que los logros de la Inteligencia Artificial en nuestros días —ya sea identificando imágenes, reconociendo el habla o ganándonos al ajedrez— están muy lejos de reproducir la inteligencia humana.

Minsky solía poner como ejemplo que cualquier chico sabe que puede tirar de un objeto con una soga, pero no empujarlo, mientras que ninguna computadora tiene este conocimiento.

La inteligencia humana va mucho más allá de la lógica que puede implementarse en una computadora. La gente usa un paraguas para evitar mojarse, decía Minsky, pero no le importa mojarse cuando va a una pileta. La valija que se usa para transportar ropa, también puede usarse para pararse sobre ella y cambiar un foquito de luz. Este es el rasgo más notable de la inteligencia humana, la capacidad de crear analogías y hacer diversos usos de las cosas, de desarrollar el sentido común y tener infinidad de maneras posibles de encarar y resolver problemas. La inteligencia humana permite entender una fábula y el doble sentido, yendo más allá del sentido literal de las palabras. Y aunque no necesariamente lleguemos a ser un gran maestro de Go, podemos aprender a jugarlo sin necesitar millones de ejemplos. Lee Sedol perdió contra AlphaGo, pero lo notable es que pudiera siquiera hacerle partida, habiendo jugado una ínfima cantidad de partidas comparado con la computadora. Esto demuestra la superioridad humana y pone el foco en lo que todavía las máquinas no han logrado: la capacidad de inteligencia general, de abstraer y transferir conocimiento. Una máquina que interpreta imágenes no puede reconocer el habla, mientras que cualquier chico de un par de años ya puede hacerlo. Para un jugador amateur de ajedrez sería relativamente sencillo aprender un juego alternativo en el que se cambian las reglas (por ejemplo, jugando a comer los peones del adversario). Sin embargo, los algoritmos de computadora, por más que venzan al mejor de los ajedrecistas, necesitan infinidad de nuevos ejemplos y entrenamiento para jugar de acuerdo a las nuevas reglas. Los avances de la Inteligencia Artificial en estos últimos años han sido realmente sorprendentes, pero la inteligencia humana es todavía muy superior a la de una máquina; no necesitamos entrenarnos con millones de ejemplos para cada tarea que resolvemos, porque somos el resultado de millones de años de evolución.

1 Imagen cortesía de Gregory Loan.

2 Un pensamiento no puede interactuar con la materia a no ser que creamos que, contando con dotes excepcionales, podamos lograr mover objetos a distancia o trasmitir pensamientos como un maestro Jedi. Estas capacidades paranormales, particularmente la telequinesis y la telepatía, no existen. El aclamado mago James Randi se dedicó a desenmascarar a aquellos que aseveraban tener tales poderes y ofreció un premio de 1 millón de dólares a quien pudiera demostrar habilidades paranormales. No es que Randi fuera multimillonario, sino que estaba seguro de que nunca nadie podría reclamar el premio. Entre otros, Randi humilló públicamente al entonces famoso "mentalista" israelí Uri Geller, que aducía poder doblar objetos con su mente o leer el pensamiento. Randi también mostró cómo un evangelista que declaraba poder adivinar y curar los males que aquejaban a su audiencia, en realidad recibía señales de radio de su mujer que le pasaba información en base a plegarias que la audiencia había provisto antes de la función. Lo realmente sorprendente es que estos charlatanes, por más expuestos que queden, puedan aún seguir haciendo fortunas a través de donaciones o la venta de sus libros.
En la Argentina hubo infinidad de embaucadores que se atribuyeron poderes psíquicos. En su cruzada para desenmascarar estos engaños, el conductor Raúl Portal mostró como dos "telépatas" venezolanos se pasaban información a través de un sencillo código que usufructuaba su habitual floreo verbal. Por ejemplo, para transmitirle a su cómplice lo que una persona había dibujado, uno de ellos decía muy rápidamente: Concéntrate, Ahora, Rápido, Ahora… Su cómplice, con los ojos vendados, no hacía más que unir las primeras letras de cada palabra para "milagrosamente" adivinar que el dibujo era una CARA.
3 Para más detalles sobre Phineas Gage, ver, por ejemplo: Malcolm Macmillan. "Phineas Gage – Unravelling the myth". *The Psychologist*; 21: 828-831; 2008.
4 El problema de Tan no era de comprensión del lenguaje, sino de articulación de un discurso fluido (tanto oral como escrito), a pesar de no presentar dificultades en el movimiento de la lengua, labios, boca, cuerdas vocales, etcétera. De hecho, podía silbar o tararear una melodía sin problemas. Poco después, en 1879, el neurólogo alemán Carl Wernicke describiría el fenómeno opuesto: pacientes con lesiones en el lóbulo temporal izquierdo que podían hablar (aunque sólo frases incoherentes) pero sin comprender el lenguaje tanto oral como escrito. El lenguaje es entonces procesado por dos áreas conectadas, lo que se conoce como el área de Wernicke, encargada de la comprensión del lenguaje, y el área de Broca, encargada de su ejecución. Lesiones en el área de Wernicke y de Broca dan lugar a dos patologías distintas llamadas "afasia sensorial" y "afasia motora". Los pacientes con afasia motora (de Broca) apenas pueden pronunciar palabras aisladas, y aquellos con afasia sensorial (de Wernicke) pueden hablar sin problemas, pero lo que dicen no tiene sentido. En algunos casos la lesión abarca ambas áreas produciendo lo que se conoce como afasia mixta.
5 H. M. fue estudiado durante 30 años por la psicóloga canadiense Brenda Milner. Para más detalles, ver el primer reporte científico que Milner escribiera junto a

Scoville sobre H. M.: William Scoville and Brenda Milner. “Loss of recent memory after bilateral hippocampal lesion”. *Journal of Neurology, Neurosurgery and Psychiatry* 20: 11-21; 1957.

6 Gilbert Ryle; *The concept of mind*; Routledge; 1949.

7 Junto a James Watson y Maurice Wilkins, Francis Crick recibiría el Premio Nobel de Medicina y Fisiología en 1962 por descubrir la estructura de doble hélice del ADN. En las últimas décadas de su vida se dedicó a estudiar el cerebro y el problema de la conciencia, mayormente junto a Christof Koch (quien fuera mi mentor en el Instituto Tecnológico de California).

8 Para más detalles del algoritmo, ver: Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton. “Imagenet classification with deep convolutional neural networks”. *Advances in neural information processing systems* 25, 1097-1105; 2012.

9 Para una recopilación de las performances de distintos algoritmos en los últimos años y cómo estas comparan con la performance humana ver: Olga Russakovsky *et al.* “ImageNet Large Scale Visual Recognition Challenge”. *International Journal of Computer Vision* 115:211-252; 2015.

10 Para más detalles, ver: Yann LeCun, Yoshua Bengio and Geoffrey Hinton. “Deep Learning”. *Nature* 521: 436-444; 2015.

11 Ver LeCun *et al.* “Deep learning”. *Nature* 521: 436-444; 2015.

12 Christof Koch. “How the computer beat the Go master”. *Scientific American*, March 19; 2016.

13 David Silver *et al.* “Mastering the game of Go with deep neural networks and tree search”. *Nature* 529, 484-489; 2016.

II. Blade Runner

¿Sienten los androides?

En *Blade Runner*, Ridley Scott retrata una Los Ángeles del futuro, dominada por grandes corporaciones de edificios imponentes y una noche perpetua con reflejos de neón y gigantescos carteles publicitarios. La acción transcurre en noviembre de 2019, un futuro lejano para cuando se estrenó la película en 1982. Sin embargo, la visión de Scott no difiere mucho del panorama que en nuestros días ofrece la noche de algún suburbio de Tokio.

Basada en un clásico de ciencia ficción escrito por Philip Dick, *¿Sueñan los androides con ovejas eléctricas?*, la trama es un *film noir* futurista en donde Rick Deckard, un cazador de androides (o "Blade Runner"), persigue a un grupo de "replicantes Nexus 6", una serie de androides idénticos a los humanos, que han entrado ilegalmente a la Tierra. En el futuro imaginado por Dick, los replicantes son usados como mano de obra esclava para trabajos peligrosos en colonias en otros planetas y tienen prohibido el ingreso a la tierra, bajo pena de muerte (que en su caso se denominaba "retiro"). Estos androides fueron genéticamente diseñados a semejanza del hombre, con la única excepción de que no tienen sentimientos. Sin embargo, tras unos años comenzaron a desarrollar sus propias respuestas emocionales volviéndose inestables y, por ese motivo, se les incorporó un mecanismo genético que les permitía vivir solo cuatro años.

La trama de la película se dispara a partir de este hecho. Sabiendo de su relativamente corta existencia, un grupo de replicantes irrumpe en la Tierra para enfrentar a su creador y exigirle más años de vida. Pero el tema de fondo es mucho más interesante. La película, como el libro de Dick, cuestiona la esencia de lo que significa ser humano y nos lleva a preguntarnos por lo que realmente diferenciaría a un hombre de un androide. Quizás lo más memorable de la película sean las últimas palabras de Roy Batty, el líder de los replicantes:

[I've seen things you people wouldn't believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the darkness at Tannhäuser Gate.
All those moments will be lost in time like tears in rain. Time to die…]

He visto cosas que ustedes humanos no creerían. Naves de ataque en llamas en la constelación de Orión. Vi rayos-C brillar en la oscuridad en la Puerta de Tannhäuser.
Todos estos momentos se perderán en el tiempo como lágrimas en la lluvia. Tiempo de morir…

Los recuerdos de Batty son aquello que lo diferencian de otros replicantes. Estos recuerdos lo hacen sentirse persona a pesar de no ser humano, justificando su afán de aferrarse a una corta vida y, eventualmente, tratar de prolongarla. Tras verlo morir, Deckard reflexiona que él en el fondo buscaba respuestas a las preguntas que todos nos hacemos: ¿de dónde venimos?, ¿hacia dónde vamos?, ¿cuánto tiempo tenemos?[14]. Tan difusa es la separación entre androides y humanos en Blade Runner, que Deckard termina enamorándose de un androide y los fanáticos de la película aún discuten si él mismo no era también un replicante.

La discusión de fondo está muy emparentada con un planteo que ha dado lugar a debates intensos en la filosofía de los últimos años: el famoso *zombi de los filósofos*. El argumento es el siguiente: imaginemos que fuera posible replicar a una persona hasta el más mínimo detalle, reproduciendo su cerebro, con todas sus neuronas y las conexiones entre estas; ¿será esta réplica un ser sintiente, o será un zombi que se comporta como un humano pero sin tener consciencia de su propia existencia?

Ilustremos esta idea en el contexto de *Blade Runner*. Imaginemos que en un futuro próximo el hombre descubre ilimitadas cantidades de uranio en un planeta cercano. La explotación de semejante reserva implicaría la solución a todos los problemas energéticos en la Tierra. Sin embargo, una persona dedicada a explotar estas minas apenas lograría sobrevivir unos pocos años, ya que los niveles de radiación en este planeta son altísimos.

La solución, entonces, sería llevar replicantes a hacer este trabajo. Después de todo, al carecer de consciencia no son más que máquinas a las cuales no tenemos ningún problema en mandar a una muerte segura. Pero para esto necesitamos que los replicantes sean zombis, necesitamos que no tengan conciencia, porque si la tuvieran serían seres sintientes a los que estaríamos condenando a una muerte injusta. Por otro lado, los replicantes (así como lo hicieron en la película) se rebelarían a tal destino y lucharían por ser tratados como humanos.

¿Es posible replicar a una persona y lograr que se comporte igual que un humano, aunque sin ser consciente? Algunos filósofos contemporáneos, como el australiano David Chalmers, argumentan que tales zombis son en principio posibles dado que: es posible concebirlos; todo lo que es concebible es posible, y consecuentemente los zombis son posibles[15]. Sin embargo, si le damos la razón a Chalmers, tendríamos que preguntarnos qué es lo que nos diferencia de estos zombis o, en otras palabras, qué es lo que nos hace conscientes.

En la serie de televisión inglesa *Humans*, estrenada en 2015, grupos de androides asisten a los humanos en trabajos rutinarios, que van desde tareas en fábricas hasta menesteres domésticos como limpiar, ordenar o preparar la comida. Estos androides son la ayuda perfecta, ya que carecen de consciencia y solo necesitan recargar su batería para seguir trabajando sin pausa, sin siquiera preguntarse qué es lo que hacen y por qué.

La trama de la serie se dispara cuando algunos de estos androides se vuelven conscientes y empiezan a plantearse su destino y la razón de su existencia. El “despertar” de los androides se da una vez que se les incorpora un complejo algoritmo. ¿Pero cómo sería ese algoritmo? ¿Qué tendría que hacer para volver consciente a un zombi?

Si nos martillamos accidentalmente un dedo gritamos de dolor, pero el dolor está dado por la activación de un grupo específico de neuronas a partir de la transmisión de información por nervios llamados *fibras C*. Un breve diálogo en la segunda parte de *Terminator* ilustra este problema. Mirando sorprendido su campera de cuero con decenas de agujeros de bala, John Connor le pregunta a Terminator

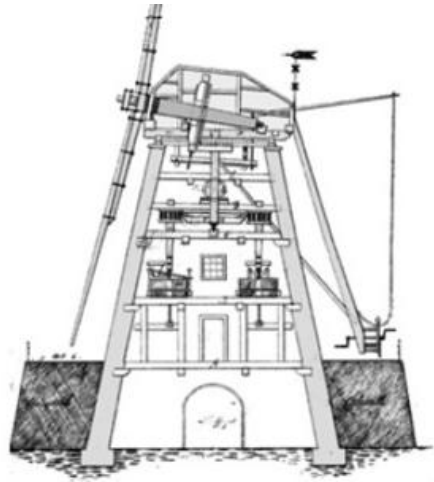
“¿Te duele cuando te disparan?”, a lo que Terminator contesta: “Detecto lesiones. Esta información puede ser llamada dolor”. En este contexto, nos preguntamos cómo la activación de neuronas y nervios específicos genera la sensación de dolor. O eventualmente: ¿por qué las neuronas que se activan al martillarnos un dedo nos hacen sentir dolor y no placer, tristeza o cosquillas? En neurociencia sabemos que al ver un cartel rojo se activan neuronas en un área de la corteza visual llamada V4[16]. Algunas de estas neuronas responderán al color rojo, otras al azul y otras al amarillo, pero esto no explica cómo se genera la sensación del rojo, la experiencia que tenemos al ver una frutilla, una rosa o la capa de un torero. Esta experiencia subjetiva es lo que se conoce como *qualia*.

Reformulando la pregunta anterior, los científicos aún discutimos cómo determinados procesos físicos y materiales (ya sea el impacto de ondas de luz en nuestra retina o la activación de un grupo de neuronas) dan lugar a experiencias subjetivas. Este es el *explanatory gap* (la brecha en nuestra capacidad de explicación), o lo que Chalmers llama *the hard problem* (el problema difícil), el problema que para él no tiene ni tendrá solución.

El pesimismo de Chalmers no es del todo infundado, después de todo estamos lidiando con nuestra propia consciencia y carecemos de un punto de vista objetivo. Podemos entender por qué una manzana cae al piso, por qué un perro saliva al escuchar una campana que precede la comida, pero ¿cómo entender justamente aquello que nos hace entender estos procesos?, ¿cómo salirnos de la visión subjetiva de nuestro propio pensamiento para poder evaluar objetivamente lo que nos pasa? Por otro lado, quizás no seamos lo suficientemente inteligentes como para entendernos a nosotros mismos. Aunque cabe aclarar que el hecho de que no podamos entender algo no implica que ese algo no exista.

Gottfried Leibniz, el brillante matemático y filósofo que desarrollara (independientemente de Newton) los principios del cálculo matemático, ya se hacía estas preguntas a principios del siglo XVIII (*Monadología*, §17):

> Por otra parte, hay que confesar que la percepción y lo que de ella depende es inexplicable por razones mecánicas, es decir, por medio de las figuras y de los movimientos. Y si se imagina que existe una máquina, cuya estructura haga pensar, sentir, tener percepción, se la podrá concebir agrandada, conservando las mismas proporciones, de tal manera que se pueda entrar en ella como si fuera un molino. Supuesto esto, se hallarán, visitándola por dentro, solo piezas que se impulsan las unas a las otras, y nunca nada con qué explicar una percepción.



Gottfried Leibniz y el mecanismo de un molino de viento, el cual usa para desafiar la posibilidad de entender cómo funciona el cerebro.

Lo mismo que afirmaba Leibniz sobre la percepción puede decirse sobre la conciencia. Pero en contraste con la visión pesimista de Leibniz y Chalmers, muchos científicos contemporáneos creen que es posible entender la conciencia, aunque este quizás continúe siendo por muchos años el gran desafío de la ciencia.

De hecho, hasta hace unas décadas nos preguntábamos por aquello que hace que algo esté vivo. ¿Qué es lo que hace que una mosca, un ombú o una ameba tengan vida y una roca no la tenga? A principios del siglo XX, el filósofo francés Henri Bergson planteó la existencia de un *elan vital* (impulso de vida), que pasa de una generación a la otra en los seres vivos. Sin embargo, hoy sabemos que tal sustancia misteriosa no existe y que lo que se transmite de una generación a otra es el ADN (ácido desoxirribonucleico). De la misma manera,

podemos preguntarnos si la conciencia es algo misterioso e inaccesible al conocimiento humano, o si algún día llegaremos a explicarla a partir de principios científicos, al igual que ocurrió con el *elan vital*.

Volvamos a *Blade Runner*. Imaginemos enfrentar al grupo de replicantes liderados por Roy Batty que demanda saber por qué no son tratados como humanos; por qué tienen negado el acceso a la Tierra y son explotados en trabajos insalubres. En principio, podríamos argumentar que ellos no son tratados como humanos porque no tienen consciencia, porque no tienen una mente. Mas al querer diferenciarnos a partir de la existencia de nuestra mente, estaríamos cayendo en la trampa del dualismo cartesiano (una trampa que es difícil evitar porque la visión dualista del cuerpo y la mente está muy arraigada en nuestro pensamiento) y, como dice Daniel Dennett, tal vez el filósofo más brillante de nuestros días, aceptar el dualismo cartesiano es darse por vencido (*Consciousness explained*, 1991, p. 37).

La mayoría de los científicos contemporáneos promulgan el *materialismo*; la idea de que la mente no es más que la actividad de neuronas[17]. Si replicamos entonces las neuronas de una persona, también replicamos su mente y tales copias no serían zombis errantes por el mundo, sino seres conscientes e indistinguibles del resto de las personas. Al abrazar el materialismo nos enfrentamos, sin embargo, al "problema difícil" que planteara Chalmers: entender cómo la actividad de neuronas puede generar la consciencia.

Vale aclarar que uno de los obstáculos más importantes en el estudio de la consciencia es un problema estrictamente semántico, ya que no tenemos una definición precisa de qué es la consciencia. Tendemos a usar vagamente la misma palabra para describir distintos procesos que están relacionados, pero que no son lo mismo. De hecho, no es lo mismo la conciencia que tenemos de nuestra propia existencia, aquella que se desvanece en el sueño, que la conciencia que tenemos de estar mirando una flor o de estar pensando en el teorema de Pitágoras.

Para evitar estos problemas, Francis Crick y Christof Koch propusieron un radical cambio de paradigma a principios de los 90[18]. Hasta ese entonces las discusiones sobre la conciencia eran de dominio

casi exclusivo de los filósofos, y la propuesta de Crick y Koch fue dejar de lado discusiones sobre definiciones y clasificaciones de la conciencia (tras varios siglos los filósofos aún no se ponen de acuerdo), dejar también de lado soluciones universales que lo expliquen todo y fomentar, en cambio, el desarrollo de paradigmas experimentales que contesten al menos distintos aspectos vinculados a la conciencia.

En otras palabras, en vez de enfocarse en el problema difícil, la idea es explorar experimentalmente problemas más sencillos relacionados con la conciencia y, una vez adquirida cierta base de conocimiento, volver entonces a la gran pregunta. Es decir, en vez de plantearnos ¿qué es la conciencia? o ¿cómo podemos ser conscientes? (algo que efectivamente nos plantearemos más adelante), pasamos a preguntarnos, por ejemplo: ¿qué pasa en el cerebro cuando reconocemos una imagen, y cuál es la diferencia cuando no la reconocemos?

Uno de los paradigmas más usados para abordar experimentalmente esta última pregunta es el de *rivalidad binocular*. La idea es muy simple. Nuestros ojos están expuestos al mismo campo visual, aunque algo desfasado, ya que están separados unos pocos centímetros. Este corrimiento de la imagen genera la sensación de ver en profundidad. Las películas 3D no hacen más que proyectar, a través de anteojos especiales, imágenes desfasadas a cada ojo y por eso vemos que las cosas parecen salirse de la pantalla. Sin embargo, desde que nacemos, ambos ojos ven siempre la misma imagen con un mínimo corrimiento.

¿Qué pasaría entonces si, usando anteojos 3D, en vez de proyectar la misma imagen desfasada a cada ojo, proyectamos dos imágenes totalmente distintas? La respuesta es que en lugar de ver una fusión de ambas imágenes, vemos una o la otra, alternando espontáneamente cada pocos segundos[19]. Este principio es ideal para testear la conciencia. La clave es que el estímulo que impacta los ojos no cambia (lo que ve cada ojo es siempre lo mismo), mientras que la sensación subjetiva, es decir, la percepción de ver una o la otra imagen, va alternándose en el tiempo.

¿Qué harán entonces las neuronas en el cerebro? ¿Responderán

siempre de la misma manera, ya que es el mismo estímulo visual, o modularán su actividad de acuerdo con la percepción que tenga el sujeto? Esta es exactamente la pregunta que se hicieron Nikos Logothetis (un brillante científico, director de un Instituto Max Planck en Tübingen) y sus colaboradores, registrando la actividad de neuronas en la corteza visual de monos. Resumiendo muchos años de trabajo en una frase, encontraron que la gran mayoría de neuronas en la corteza visual primaria (la primera área en la corteza cerebral que procesa los estímulos visuales) no cambiaban su actividad, mientras que la mayoría de las neuronas en áreas visuales superiores sí lo hacían, siguiendo la percepción subjetiva de los monos[20].

Los trabajos de Logothetis y sus colegas dan entonces una sólida evidencia experimental que va a las raíces mismas del tan mentado problema difícil de la consciencia. ¿Cómo puede entonces un estímulo material dar lugar a una sensación subjetiva? Muy fácil: neuronas en áreas visuales primarias representan las características básicas del estímulo, y tras sucesivas etapas de procesamiento se genera una representación de la sensación subjetiva en áreas superiores. Los distintos niveles de procesamiento se encargan de formar una representación cada vez más abstracta para extraer un significado del estímulo[21]. No es casual que en el capítulo anterior mencionáramos que los sucesivos niveles de las redes profundas de Hinton generan una representación análoga de las imágenes que analizan. El diseño de tales redes está justamente basado en los principios de procesamiento visual usados por el cerebro.

Las áreas visuales superiores proyectan al hipocampo, aquella estructura con forma de caballito de mar que fuera extraída en el paciente H. M. para tratar de curarlo de la epilepsia. Vimos que a partir de la cirugía H. M. perdió la capacidad de formar nuevas memorias. Basados principalmente en la información provista por H. M. y en posteriores estudios con lesiones en animales[22], hoy nadie duda del rol clave que tiene el hipocampo en la memoria. Pero, sorprendentemente, la cirugía que tuviera resultados desastrosos para H. M., en la actualidad se realiza frecuentemente y con chances bastante altas de curar al paciente[23].

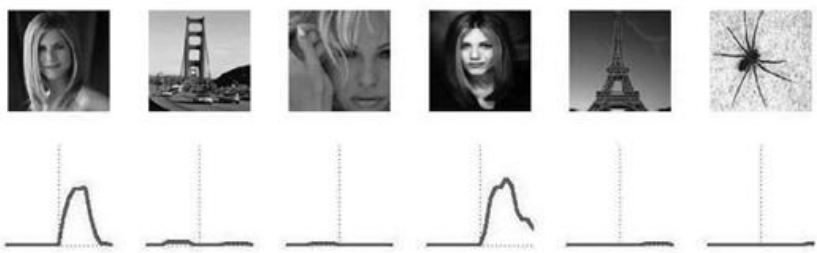
La mayoría de pacientes con epilepsia pueden ser tratados con medicación, pero en alrededor de un 20% de los casos la medicación no logra controlar las crisis y estos pacientes son en principio candidatos a una resección quirúrgica del *foco epiléptico*. En muchos casos, tales resecciones incluyen al hipocampo porque suele estar involucrado en el origen de las crisis epilépticas. Sin embargo, al hacer esto ¿no estaríamos condenando a los pacientes a terminar amnésicos como H. M.? Por supuesto que no. La gran diferencia es que se extrae solo uno de los hipocampos, el izquierdo o el derecho (y algunas áreas aledañas), según el caso, pero nunca ambos.

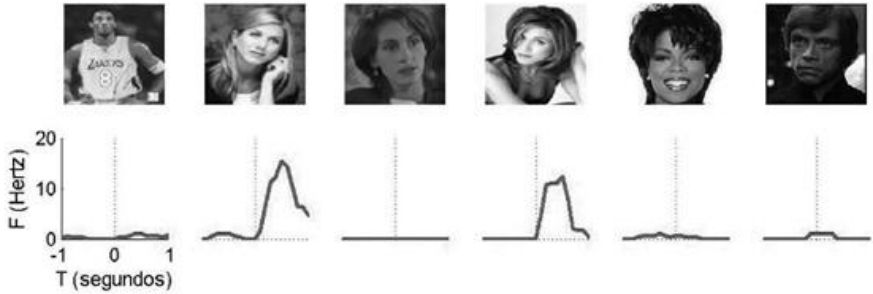
Ahora bien, si se evalúa extraer quirúrgicamente un área del cerebro de un paciente, es vital asegurarse de que es efectivamente el área que inicia las crisis epilépticas. Por este motivo, en muchos casos se procede a la implantación de electrodos en el cerebro, para registrar su actividad durante las crisis epilépticas y así poder determinar con la mayor precisión posible la localización del foco.

Para evitar cualquier confusión, reitero que el motivo por el cual se colocan estos electrodos (así como la decisión de cuántos electrodos se colocan en cada paciente y en qué áreas) es exclusivamente clínico; el objetivo es curar al paciente. Aunque, por otro lado, tal procedimiento permite estudiar el comportamiento de neuronas en el cerebro humano[24]. La posibilidad única de hacer este tipo de registros fue lo que me convencería de hacer un posdoctorado en el Instituto Tecnológico de California (Caltech) y lo que me llevaría al descubrimiento que cambiaría radicalmente mi carrera científica.

Implantación de electrodos intracraneales en un paciente con epilepsia.

En los primeros experimentos les mostramos a los pacientes fotos de actores, políticos, deportistas, lugares famosos, etcétera, mientras registrábamos la actividad de decenas de neuronas[25]. La idea era ver si alguna de estas neuronas respondía a alguna de las fotos. Dejando de lado infinidad de detalles técnicos, el resultado superó por lejos incluso las expectativas más optimistas que tenía al mudarme a Los Ángeles para desarrollar estos experimentos: encontramos neuronas en el hipocampo humano que respondían a conceptos[26]. La más famosa de estas neuronas fue una que respondió a distintas fotos de Jennifer Aniston y no a fotos de muchas otras personas (incluyendo otras actrices como Julia Roberts, Pamela Anderson, etcétera).





Respuestas de una neurona en el hipocampo que disparó solo a distintas fotos de Jennifer Aniston (por cuestiones de espacio, se muestran solo cuatro de las siete fotos de Jennifer Aniston y solo ocho de las otras ochenta fotos usadas). Los trazos gruesos muestran la respuesta promedio a seis presentaciones de cada foto, mostrada a partir del tiempo cero.

Asimismo, encontramos otra neurona que solo respondió a distintas fotos de Halle Berry, otra que lo hizo con las de Oprah Winfrey, otra que respondió a Luke Skywalker y decenas de otros ejemplos[27]. El hecho de que estas neuronas se activaran con distintas fotos de una persona (y no con las de otras), implica que responden a la persona, al concepto, y no a detalles particulares de una u otra foto. Esto fue lo más sorprendente para mis colegas (y por supuesto para mí mismo): el increíble grado de abstracción y de representación de conceptos de estas neuronas. Es más, las neuronas también se activaban cuando mostrábamos el nombre escrito de la persona a la cual respondían (y no cuando mostrábamos otros nombres) e incluso cuando el nombre era pronunciado por una voz sintetizada en la computadora[28].

Di un rodeo enorme para, ahora sí, poder explicar un par de experimentos que están muy ligados a la discusión anterior sobre la conciencia. Supongamos que le mostramos al paciente una foto de Jennifer Aniston, pero en vez de dejarla en la pantalla por un segundo, como en el experimento inicial, la mostramos muy brevemente (durante menos de una décima de segundo, para ser más concretos). La idea es que, al mostrar la foto tan rápido, el paciente a veces reconoce a la persona y otras veces no. Supongamos también que hacemos esto mientras registramos la actividad de la neurona que responde a Jennifer Aniston. ¿Qué hará entonces esta neurona?

¿Responderá todas las veces que mostramos la foto (después de todo, el estímulo visual es exactamente el mismo) o solo cuando el paciente reconoce a Aniston? Esta es justamente la idea del experimento que hicimos hace unos años, donde comprobamos que si el paciente reconocía la foto, aunque esta fuera mostrada por una mínima fracción de segundo, la neurona respondía igual que si la foto quedaba en la pantalla durante mucho más tiempo. Por el contrario, si el paciente alegaba no haber visto la foto, la neurona permanecía callada, como si no hubiéramos mostrado nada. En otras palabras, estas neuronas reflejan la percepción consciente del paciente y no el estímulo visual, el cual era siempre el mismo[29].

En un trabajo posterior continuamos explorando la misma idea pero con otro diseño experimental. En este caso, en vez de mostrar rápidamente una foto para que el paciente a veces la vea y a veces no, usamos distintas transformaciones entre dos caras. Por ejemplo, partiendo de una foto de Angelina Jolie y otra de Halle Berry, creamos distintas transformaciones entre ambas caras y evaluamos cuál de ellas daba la respuesta más ambigua; es decir, cuál de estas transformaciones el paciente a veces veía como una persona (Jolie) y otras veces como la otra (Berry). Entonces, si registramos una neurona que responde a Halle Berry, ¿qué hará esta neurona cuando mostremos la foto ambigua entre Berry y Jolie? ¿Responderá siempre de la misma manera, ya que la foto es la misma, o responderá solo cuando el paciente dice que es Halle Berry?

Esto último es lo que encontramos: las neuronas disparaban solo cuando los pacientes decían que la imagen transformada correspondía a una persona (de hecho, repondrían igual que al mostrar la foto original) y no disparaban si al ver la misma imagen decían que correspondía a la otra[30].

Transformaciones usadas para testear las respuestas de las neuronas en el hipocampo. La foto de la izquierda corresponde a Angelina Jolie, la de la derecha a Halle Berry y la del medio es una transformación ambigua entre ambas, que el paciente a veces identificaba como Jolie y otras veces como Berry.

Resumiendo los resultados de los experimentos anteriores, en la corteza visual primaria existe una representación del estímulo material, mientras que en áreas visuales superiores hay una representación de la percepción subjetiva del sujeto. Las áreas visuales superiores, a su vez, envían información al hipocampo, en donde la representación es mucho más abstracta y ligada a la consciencia del sujeto. No importa si lo que mostramos es la foto de una persona, su nombre escrito o dicho por una computadora; no importa si apenas mostramos un destello de una foto o si usamos una imagen transformada; mientras el sujeto reconozca a la persona, sea a partir del estímulo que sea, las correspondientes neuronas del hipocampo —las neuronas de concepto— van a activarse.

¿Están, entonces, las neuronas del hipocampo involucradas en la consciencia? La pregunta tiene trampa ya que es muy vaga. Depende de lo que entendamos por consciencia. Si por consciencia nos referimos al hecho de reconocer algo, es decir, de ser consciente de estar viendo a Halle Berry y no a Angelina Jolie, la respuesta es un rotundo no. ¿Por qué? Porque al paciente H. M. se le extrajeron ambos

hipocampos, pero igual podía reconocer a su madre, al cirujano que lo operó, la casa donde vivía y otras personas y cosas que conocía de antes de la cirugía. H. M. no podía formar nuevas memorias, pero su percepción era normal[31]. Aunque, por otro lado, no es lo mismo el proceso de reconocer una foto de Halle Berry, que traer al consciente una avalancha de recuerdos relacionados con ella ("recuerdo haberla visto en una película de James Bond", "recuerdo haber encontrado una neurona que respondía a ella", "recuerdo haber usado imágenes en las que la transformaba en Angelina Jolie", etcétera). Y estos recuerdos que traigo al consciente pueden ser muy diversos: al ver una foto de un jazmín recuerdo su olor, la textura de sus pétalos, la planta de jazmín del país en la casa de mi abuela en Chacabuco. Todos esos recuerdos forman parte de mi experiencia subjetiva al ver un jazmín. La suma de todas estas sensaciones y recuerdos es lo que llamamos *qualia*.

La conciencia involucra entonces al menos dos procesos distintos (y probablemente muchos más): el de reconocer una persona, por un lado, y el de traer al consciente la información relacionada con esa persona, por el otro[32]. ¿Y qué pasaría si no tuviéramos hipocampo, si no tuviéramos las neuronas que acabo de describir? Eventualmente podríamos reconocer objetos, animales y personas, pero probablemente no recordaríamos las experiencias relacionadas con ellos. Tal era el estado de H. M. tras la cirugía, de permanente confusión, de vivir en un eterno presente sin acceso al pasado. De hecho, H. M. solía decir que "cada día es único en sí mismo, cualquiera sean las alegrías o las penas que haya tenido".

Leonard, el personaje principal de la película *Memento*, ofrece una espectacular caracterización de H. M. y lo que implica vivir sin la capacidad de formar nuevas memorias. Para recordar algo relacionado a una persona, Leonard escribía notas en una foto polaroid de dicha persona. La película gira en torno a las dramáticas decisiones que toma a partir de estas notas, como si constituyeran su única fuente de memoria duradera. Volviendo a H. M., Brenda Milner, quien lo estudiara durante 30 años[33], describía su estado como el de alguien que está continuamente como empezando a entender sus

circunstancias, pero sin llegar a comprender acabadamente lo que sucede, porque no puede poner su situación presente en el contexto de su pasado inmediato. H. M. decía al respecto que su vida era como el despertar de un sueño.

La mayoría de las veces despertamos en nuestra cama y todo nos resulta familiar, pero si pasamos la noche en un hotel, al despertar solemos tener ese breve instante de confusión en el que no entendemos dónde estamos, ni qué hacemos allí. Sin embargo, nuestra memoria rápidamente nos rescata recordándonos las circunstancias por las cuales estamos en ese lugar y todo vuelve a tener sentido. En el caso de H. M. ese momento de confusión se prolongaba interminablemente porque no tenía memoria de dónde se encontraba y qué hacía en donde fuera que estuviere. Ilustrando brillantemente esta idea, la primera escena de *Memento* comienza con la voz en *off* de Leonard al despertar en un lugar desconocido:

> ¿Dónde estoy? Una habitación de un motel. Acabo de despertar y estoy… en una habitación de un motel. Acá hay una llave. Siento que quizás es la primera vez que estoy aquí pero… quizás estoy aquí hace una semana… tres meses… Es difícil de decir… no sé. Es sólo una habitación anónima.

El caso de H. M. quizás sea lo más cercano al zombi de los filósofos. Ya establecimos que tales zombis en realidad no podrían existir; si replicáramos el cerebro de un humano, también replicaríamos su consciencia. Pero si en el proceso de clonación hubiera una falla al reproducir el hipocampo, el ser que crearíamos se comportaría en principio como otros humanos (a pesar de su estado de confusión permanente, H. M. podía reconocer cosas, entablar una conversación y hasta tenía un coeficiente intelectual superior a la media), pero tendría problemas de memoria y una muy limitada consciencia de sí mismo y de sus circunstancias (y nótese que no tomo a la consciencia como algo que existe o no existe, sino como algo que puede darse en distintos grados). Curiosamente, cuando H. M. tenía 55 años no se reconoció en una foto que había sido tomada 15 años antes[34]. Tampoco era consciente del paso del tiempo y de los acontecimientos.

Por ejemplo, en abril de 1955 (un año y medio después de la cirugía), apenas tenía conciencia de haber sido operado, y estimó que la fecha era marzo de 1953 y que tenía 27 años, cuando ya había cumplido 29.

Vimos cómo el cerebro genera sensaciones subjetivas a partir de una representación de los estímulos materiales. También vimos que, al menos en parte, ya entendemos cómo se genera aquello que llamamos consciencia. Pareciera que hemos contestado la gran pregunta, que hemos resuelto el problema difícil. Pero a pesar de haber hecho grandes avances en los últimos años, el problema aparentemente sigue en pie. Después de todo, aún nos queda explicar cómo las neuronas en el hipocampo, o en las áreas visuales superiores, generan nuestras sensaciones subjetivas. Tendríamos que explicar, por ejemplo, por qué somos conscientes de la actividad de estas neuronas y no de la actividad de otras tantas que regulan la secreción de hormonas o el diámetro de las pupilas. ¿Qué es lo que nos hace conscientes de la actividad de unas y no de las otras? ¿Qué es, en el fondo, lo que da lugar a sentir algo? Estas preguntas continúan generando grandes debates y volveremos a plantearlas más adelante en este libro.

En el capítulo anterior describimos los sorprendentes avances de la Inteligencia Artificial en los últimos años, sobre todo a partir del uso de redes profundas. Argumentamos que la capacidad de consciencia no está limitada a seres orgánicos, sino que también podría, en principio, darse en circuitos y algoritmos lo suficientemente sofisticados. Pero ¿cómo saberlo? ¿Cómo podríamos testear si una máquina es consciente, si es efectivamente capaz de pensar por sí misma o si solo se limita a dar sofisticadas respuestas basadas en complejos algoritmos?

En *Blade Runner*, Rick Deckard examina a potenciales androides usando una Voigt-Kampff machine, una máquina que puede medir respuestas fisiológicas inconscientes (como la dilatación de las pupilas), mientras hace distintos tipos de preguntas con alta carga emocional. La idea es ver si la persona interrogada puede mantener una conversación coherente sobre su pasado y mostrar respuestas fisiológicas que, al carecer de sentimientos y emociones, los replicantes no tienen. Nuevas generaciones de replicantes podrían, sin

embargo, tener incorporada la capacidad de reproducir tales respuestas. De hecho, algunos androides de hoy, como los modelos Geminoid o Actroid, ya se acercan bastante a reproducir los gestos humanos. Sin embargo, lo más difícil para estos androides sería poder mantener una interacción y una conversación coherente, sabiendo cuándo y cómo reaccionar. Esta es justamente la idea en la que se basa el test de Turing, concebido por el matemático inglés Alan Turing en 1950.

Turing planteó que preguntarse si una máquina es capaz de pensar es análogo a preguntarse si una máquina puede replicar el comportamiento humano. Concretamente, propuso lo que llamó *the imitation game* (el juego de la imitación), en el que un examinador interactúa al mismo tiempo con una persona real y con una computadora en otra habitación y trata de determinar quién es la máquina y quién el humano[35].

La interacción se da a través de un monitor y un teclado, para dejar de lado detalles técnicos como el tono de la voz, la apariencia del robot, etcétera. Si a partir de este chat *online* el examinador no puede distinguir quién es quién, la máquina ha pasado el test.

Desde la década del 90 el test de Turing es implementado en el (desprestigiado) Loebner Price, una competencia anual que ofrece una recompensa de un par de miles de dólares al algoritmo que mejor replique una conversación humana. Tan desprestigiada está esta competencia (para muchos, un burdo efecto de publicidad sin el mínimo sostén científico) que Marvin Minsky, el padre de la Inteligencia Artificial de quien habláramos en el capítulo anterior, alguna vez ofreció irónicamente una compensación monetaria a quien lograra abolir este premio.

Más allá de estas críticas, el Loebner Price ha sido un pintoresco fórum en donde se han presentado distintos algoritmos de interacción con usuarios, llamados *chatbots* (abreviatura de *chatting robots*, robots conversadores). Por ejemplo Mitsuku[36], uno de los *chatbots* que ganó esta competencia, ofrece una conversación amena y razonablemente coherente, mientras las preguntas sean sencillas (y convenientemente tiende a cambiar el tema de la conversación si no entiende la

pregunta)[37]. Principios como los implementados en los *chatbots* son cada día más usados en aplicaciones como Siri, que permite interactuar verbalmente con tablets como iPad, ofreciendo el servicio de un (por ahora) muy limitado aunque eficiente asistente que maneja el email, calendario, búsquedas en Internet y otras tareas básicas.

Volviendo al test de Turing, la idea parecería tener sentido ya que podemos imaginar infinidad de preguntas con las que podríamos detectar, como Rick Deckard, que estamos interactuando con una máquina. Pero en la práctica los resultados son controversiales, porque están muy influenciados por la eficiencia del examinador en formular preguntas adecuadas para detectar al *chatbot*. Sin embargo, sería injusto endilgarle estas críticas a Turing. Después de todo, él buscaba proponer una manera práctica de determinar si una máquina puede pensar, evitando entrar en discusiones interminables sobre qué se entiende por pensamiento. Muy probablemente, Turing mismo aceptaría que en nuestros días existen otras maneras más eficientes de encarar esta pregunta. En el capítulo anterior vimos que algoritmos basados en redes profundas son capaces de reconocer y hasta interpretar el contenido de distintas fotos. Este logro es mucho más relevante que el hecho de si un *chatbot* llega a confundir al evaluador con quien interactúa.

En general, los *chatbots* no pueden responder a preguntas complejas o con doble sentido. Sin embargo, en febrero de 2011 una supercomputadora de IBM llamada Watson pulverizó a dos campeones históricos de *Jeopardy*, un juego de preguntas y respuestas de la televisión norteamericana. Lo sorprendente del caso es que en este juego las preguntas no son formuladas explícitamente (del tipo: ¿Quién es…? ¿Qué es…? etcétera), sino que están implícitas en textos relativamente intrincados que usan el doble sentido e información en distintos contextos. El solo hecho de entender el significado de las preguntas es en sí mismo un gran desafío para cualquier computadora, pero Watson no solo comprendía las preguntas, sino que daba con la respuesta correcta (usando su base de memoria y sin estar conectada a Internet) más rápido que los otros dos participantes, hasta entonces imbatibles.

Otra conocida demostración de los logros de la Inteligencia Artificial ocurrió en 2012. Rick Rashid, quien fuera director de Microsoft Research, dio una conferencia en China sobre el reconocimiento automático del habla. Mientras daba la conferencia, un algoritmo de reconocimiento basado en redes profundas escribía todo lo que iba diciendo en una pantalla gigante. Pero esto era solo el principio. Luego de explicar los problemas técnicos que involucran el reconocimiento del habla y los últimos avances en el área, Rashid pasó a hablar de los progresos de Microsoft en la traducción automática de textos. A continuación, explicó cómo era posible, primero, que un algoritmo reconociera lo que estaba diciendo (como se mostraba en la pantalla); segundo, que otro algoritmo lo tradujera al chino (como comenzó a mostrarse en otra pantalla), y finalmente, cómo era también posible que un tercer algoritmo pronunciara esas palabras.

Ante el asombro de todo el auditorio, Rashid terminó su conferencia hablando en chino, a través de una voz sintetizada que (con su misma voz) traducía sus palabras ni bien las pronunciaba[38]. Menciono a propósito esta demostración porque es la ilustración perfecta de una discusión que se viene dando desde hace años en la filosofía: el famoso argumento de la *habitación china*.

Amén de las ya mencionadas limitaciones del test de Turing, una crítica mucho más substancial a la Inteligencia Artificial fue hecha por el filósofo John Searle, quien argumentó que el test de Turing en realidad nunca podría afirmar si una máquina es capaz de pensar. Para ilustrar este punto, Searle propuso imaginar a una persona en una habitación con un gran manual de instrucciones acerca de cómo operar símbolos chinos, a quien, desde afuera, le pasan tarjetas con preguntas en chino, que no entiende, pero que es capaz de contestar siguiendo las instrucciones del manual.

La conclusión de Searle es que esta persona aparentaría entender chino a pesar de no hablar una palabra y pasaría el test de Turing. Searle utiliza este argumento para negar la posibilidad de que una máquina sea capaz de pensar, pues, según él, sería como el caso de la persona en la habitación china, la cual solo puede seguir reglas pero sin entender su contenido. Al enterarse del logro de Watson en

Jeopardy, Searle felicitó a los científicos de IBM, pero concluyó que Watson ni siquiera se enteró de que ganó. Por cierto, no todos comparten la visión pesimista de Searle sobre la Inteligencia Artificial[39]. Después de todo, ¿qué es lo que hace que nosotros podamos entender y una máquina no? ¿Cuál sería la diferencia, y por qué tal diferencia no podría implementarse en una máquina?

La gran mayoría de la comunidad científica está de acuerdo en que la pregunta de cómo determinar si una máquina es capaz de pensar y ser consciente aún no tiene respuesta. Describimos el test de Turing, pero también las críticas que se le hacen tanto a la idea subyacente (la discusión sobre la habitación china), como a su implementación en la práctica (las críticas al Loebner Price). En mi opinión, la persona de la habitación china demostraría entender chino si mostrara que es capaz de generalizar; es decir, si pudiera contestar al menos algunas preguntas en esa lengua sin seguir las reglas del manual de instrucciones.

Por supuesto, Searle argumentaría que este manual podría cubrir todas las preguntas posibles (y que no tenemos acceso a ver si la persona consulta el manual o no), pero esto es distinto al planteo que hiciéramos sobre la Inteligencia Artificial. Una red profunda desarrollada para reconocer imágenes no está programada, o, mejor dicho, entrenada, para jugar al ajedrez. Entonces podríamos afirmar que una máquina es capaz de pensar si puede resolver problemas para los cuales no fue explícitamente programada (o entrenada con infinidad de ejemplos), a partir de generalizar y transferir conocimiento.

Cerramos el capítulo anterior describiendo las ideas de Marvin Minsky sobre la capacidad, hasta ahora exclusivamente humana, de generalizar y utilizar el sentido común. Minsky resaltaba la importancia del uso de analogías en el pensamiento humano, en contraste con el funcionamiento de las máquinas, el cual está basado en la implementación de reglas lógicas (si pasa A, hago B). Justamente hacer analogías es la base para transferir conocimiento, para aplicar lo aprendido en una tarea a otra completamente distinta. También comentábamos en el capítulo anterior lo fácil que es para un jugador

amateur de ajedrez adaptarse a cambios de reglas, por ejemplo, jugando a comer los peones de su adversario. El jugador entiende la consigna y transfiere conocimiento sin necesitar de mucha práctica, mientras que una computadora necesita ser entrenada de nuevo desde cero.

La capacidad humana de hacer analogías y generalizar se basa en extraer un sentido de las cosas. La extracción de un sentido depende, a su vez, de poder abstraer, de quedarse con lo esencial y descartar el resto. Al montar un camello, transfiero conocimiento que adquirí montando un caballo y me voy adaptando a las peculiaridades del camello. Pero no importa si aprendí con un caballo blanco, negro o azabache. En este contexto, dejo de lado todos esos detalles y extraigo la información de cómo me balanceaba en el caballo para aplicarla al camello. Esta es también la base de nuestra creatividad, de hacer analogías y asociaciones entre hechos que aparentemente no tienen conexión. Por ejemplo, la genialidad de Newton no fue dar con la fórmula de la ley de gravedad, sino darse cuenta de que cosas tan dispares, como el hecho de que caiga una manzana y de que la Luna se mantenga en el cielo, responden al mismo fenómeno. No importa si la manzana es grande o chica, roja o verde, como tampoco importa si la Luna está en el zenit, o si es menguante o creciente. A este fin, solo importa que la manzana es un objeto estático y que por eso cae, mientras que la Luna es un objeto en órbita y, por ende, continua su rumbo.

Comparado con la capacidad de almacenamiento de información que posee una computadora, recordamos muy poco. Más que llevar un registro fehaciente de todo, nuestro cerebro busca entender. Los procesos de abstracción y atribución de significado no son nada fáciles de implementar en una computadora (ya que no es obvio decidir cuál información usar y cuál dejar de lado) y a su vez dan lugar a la construcción de analogías, a la transferencia de conocimientos y al desarrollo de la inteligencia general. Por otro lado, el proceso de abstracción genera representaciones internas y subjetivas. Esto es justamente lo que codifican las neuronas de concepto; abstracciones, conceptos que están muy ligados a nuestra consciencia, ya que estas

neuronas disparan solo cuando el sujeto es consciente de ver algo. En otras palabras, la carencia de estos procesos en una computadora es lo que hace que no tengan inteligencia general y las representaciones internas que, me atrevo a sugerir, tal vez sean la piedra angular de nuestro pensamiento y nuestra consciencia.

14 Estos pensamientos, plasmados con una voz en *off* en la versión original de la película, fueron eventualmente eliminados en ediciones posteriores. Al parecer fueron impuestos por los directivos de la compañía cinematográfica para que la trama fuese más fácil de entender, algo que Ridley Scott luego creyó innecesario.

15 David Chalmers; *The Conscious Mind: In Search of a Fundamental Theory*; Oxford University Press; Nueva York; 1996.

16 Semir Zeki. "Parallel Processing, Asynchronous Perception, and a Distributed System of Consciousness in Vision". *The Neuroscientist* 4: 365-372; 1998.

17 Para no complicar las cosas, simplifico bastante al hacer esta afirmación y no entro en distinciones filosóficas más refinadas. En el fondo, lo que quiero decir es que casi nadie en la ciencia de hoy en día sostiene el dualismo cartesiano y la idea de que la mente es algo que va más allá de la actividad de neuronas.

18 Francis Crick and Christof Koch. "The problem of consciousness". *Scientific American* 267 (3), 153-159; 1992.

19 Para más detalles, ver: Randolph Blake and Nikos Logothetis. "Visual competition". *Nature Reviews Neuroscience* 3: 13-21; 2002.
Con Mariano Molina, un gran pintor y amigo que estuvo un año en mi laboratorio, usamos este principio para crear cuadros (por supuesto realizados por Mariano) que al ser vistos con anteojos 3D cambian cada pocos segundos. Lo sorprendente es que los cuadros no son más que telas en la pared, objetos estáticos que toman movimiento cuando se los mira con los anteojos 3D. Estos cuadros, junto a muchos otros, fueron exhibidos en Inglaterra. Para más detalles, ver: http://projects.beyondtext.ac.uk/perceptionandwellbeing-fo/

20 Para más detalles, ver por ejemplo: Nikos Logothetis. "Single units and conscious vision". *Philosophical Transactions of The Royal Society* B 353: 1801-1818; 1998.

21 Queda, por supuesto, muchísimo por explicar sobre cómo funciona este proceso de abstracción y de extracción de un significado, pero no me extiendo en más detalles porque este fue el tema recurrente que desarrollara en un libro anterior (*Qué es la memoria*, Paidós, 2015).

22 Para más detalles, ver por ejemplo: "Larry Squire and Stuart Zola-Morgan. The medial temporal lobe memory system". *Science* 253, 1380-1386; 1991.

23 Aunque lamentablemente no todos los casos pueden ser tratados quirúrgicamente, ya que esto depende de la localización del foco epiléptico, el tipo de epilepsia y muchos otros factores. Al no ser médico, sería poco ético que me pusiera a discutir más en detalle sobre aspectos clínicos relacionados a la cirugía de la epilepsia. Mi objetivo al mencionar este tema es explicar el motivo por el cual se implantan electrodos en el cerebro de pacientes con epilepsia.

24 Para más detalles sobre los electrodos, el proceso de registro de neuronas y demás detalles técnicos, ver, por ejemplo: Hernan Rey, Matias Ison, Carlos Pedreira, Antonio Valentin, Gonzalo Alarcon, Richard Selway, Mark Richardson and Rodrigo Quian Quiroga, R. "Single-cell recordings in the human medial temporal lobe". *Journal of Anatomy* 227: 394-408; 2015.

25 En estos experimentos también participaron Gabriel Kreiman, Leila Reddy, Alexander Kraskov y Florian Mormann, Christof Koch e Itzhak Fried, el neurocirujano que estableciera este tipo de registros en UCLA.

26 Rodrigo Quian Quiroga, Leila Reddy, Gabriel Kreiman, Christof Koch and Itzhak Fried. "Invariant visual representation by single neurons in the human brain". *Nature* 435, 1102-1107; 2005.

27 Para más ejemplos ver: "Rodrigo Quian Quiroga, Alexander Kraskov, Christof Koch and Itzhak Fried. Explicit encoding of multimodal percepts by single neurons in the human brain". *Current Biology*. 19: 1308-1313; 2009.

28 Para un resumen de los resultados que obtuvimos con estas neuronas ver: Rodrigo Quian Quiroga. "Concept cells: the building blocks of declarative memory functions". *Nature Reviews Neuroscience* 13: 587-597; 2012.

29 Los resultados de este experimento están descriptos en: Rodrigo Quian Quiroga, Roy Mukamel, Eve Isham, Raphael Malach and Itzhak Fried. "Human single-neuron responses at the threshold of conscious recognition". *Proceedings of the National Academy of Sciences* 105: 3599-3604; 2008.

30 Los resultados de este experimento están descriptos en: Rodrigo Quian Quiroga, Alexander Kraskov, Florian Mormann, Itzhak Fried and Christof Koch. "Single-cell responses to face adaptation in the human medial temporal lobe". *Neuron* 84: 363-369; 2014.

31 El capítulo 4 de *Borges y la memoria*, op. cit. está dedicado a la descripción de H. M. Por otro lado, el primer reporte sobre el caso de H. M. es: William Scoville and Brenda Milner. "Loss of recent memory after bilateral hippocampal lesion". *Journal of Neurology, Neurosurgery and Psychiatry* 20: 11-21; 1957.
Una descripción concisa de su aporte a la ciencia está dada en: Larry Squire. "The legacy of patient H. M. for neuroscience". *Neuron* 61: 6-9; 2009.

32 Esto es lo que podríamos llamar *perceptual* y *contextual awareness*. Para más detalles, ver:

- Joaquin Navajas, Hernan Rey and Rodrigo Quian Quiroga. "Perceptual and contextual awareness: methodological considerations in the search for the neural correlates of consciousness". *Front Psychol* 5, article 959; 2014.

- Rodrigo Quian Quiroga. “Neuronal codes for visual perception and memory”. *Neuropsychologia* 83: 227-241; 2016.

33 Tras la muerte de H. M., en el 2008, Brenda Milner adujo tener una sensación extraña, la de haber perdido a un gran amigo, a pesar de que tras décadas de interacción constante ese amigo nunca llegara a reconocerla.

34 En “El marinero perdido” (cuento que es parte del libro *El hombre que confundió a su mujer con un sombrero*), Oliver Sacks hace una descripción dramática de la confusión y el horror que le generó a un paciente con amnesia el ver su rostro en el espejo, creyendo tener 19 años, cuando ya estaba cerca de cumplir 50.

35 Turing propuso su famoso test en: Alan Turing. “Computing machinery and intelligence”. *Mind* LIX: 433-460; 1950.

36 www.mitsuku.com

37 Salvando las diferencias, la interacción con estos *chatbots*, recuerdan un poco a la película *Ella*, en la que el protagonista termina enamorándose de una aplicación de computadora con la que interactúa.

38 La conferencia puede verse en YouTube, donde ya tiene más de un millón de visitas: https://www.youtube.com/watch?v=Nu-nlQqFCKg

39 Para una discusión crítica véase el artículo original de Searle y el comentario subsiguiente por diversos autores en: John Searle. “Minds, brains and programs”. *Behavioral and Brain Sciences* 3: 417-457; 1980.

III. El planeta de los simios

Consciencia animal

Un par de meses antes que Kubrick deslumbrara con *2001: Odisea del espacio*, se estrenaba *El planeta de los simios*, basada en *La planète des singes*, del novelista francés Pierre Boulle.

No es un dato menor que ambas películas sean de 1968, pues resalta la magnitud de *2001* comparada con una de las películas de ciencia ficción más taquilleras de su época. Basta contrastar las escenas de la rutina de Dave Bowman en la estación espacial de *2001* con aquellas en la nave de *El planeta de los simios* para notar la abismal diferencia entre ambas producciones. Pero esta comparación quizás sea injusta, porque *2001* podría ser estrenada mañana mismo, 50 años después, y aún sería considerada una película de vanguardia, con una estética audiovisual y una profundidad de contenido que no dejan de sorprender. Aunque esta valoración no tiene nada que ver con la recepción que tuvieron ambas películas cuando se estrenaron.

Las críticas de *2001* fueron variadas, debido a un argumento extremadamente complejo, sus interminables escenas en el espacio y una casi total carencia de diálogos. En contraste, *El planeta de los simios* fue un éxito inmediato, sustentada en una trama lineal que tocaba temas de gran actualidad en aquel entonces: el auge de la guerra fría y la posibilidad del total exterminio de la raza humana (a pocos años de la crisis de misiles de Cuba, la cual estuvo a punto de desencadenar un conflicto nuclear entre Estados Unidos y la Unión Soviética), los problemas de integración racial, justo en el año del asesinato de Martin Luther King, y la controversia que desata la experimentación en animales.

En *El planeta de los simios*, un grupo de astronautas en estado de hibernación viajan a un planeta desconocido. Debido a la velocidad de la nave, cercana a la velocidad de la luz, el tiempo pasa mucho más lento, y mientras que para los tripulantes el viaje dura un año y medio, en la Tierra han pasado más de 2000 años. Una falla los lleva a

aterrizar en un planeta al parecer desierto, en donde tras largo andar descubren a un grupo de hombres subdesarrollados, sin siquiera la capacidad del habla, que son dominados por una raza de simios mucho más avanzada. Taylor, el líder de los astronautas, es capturado y estudiado por dos simios científicos, asombrados por el interés que este muestra en tratar de comunicarse (a pesar de no poder hablar debido a una herida en su garganta al ser apresado). Probablemente la escena más lograda sea el final de la película, donde Taylor, vagando errante luego de escapar, descubre las ruinas de la Estatua de la Libertad y se da cuenta de que estuvo todo el tiempo en la Tierra, ahora dominada por los simios.

En ese futuro lejano, la raza humana ha sido brutalmente relegada tras haber sido casi exterminada por una guerra del pasado. El ejercicio que propone la película de invertir la escala evolutiva nos lleva a ponernos en la piel de especie dominada y a reflexionar sobre nuestra interacción con los animales, ya que estos muy probablemente también sean seres pensantes y conscientes como nosotros, aunque con un grado inferior de desarrollo.

Anteriormente discutimos los avances y límites de la Inteligencia Artificial y la posibilidad de pensamiento y de consciencia de máquinas y androides. En este contexto, *El planeta de los simios* es el disparador de una discusión mucho más real y cercana. Ya no hablamos de sofisticados algoritmos o circuitos que eventualmente podríamos llegar a desarrollar, sino de seres que comparten la Tierra con nosotros. ¿Podrán pensar? ¿Serán conscientes de su propia existencia? ¿Temerán su muerte? Dicho en otras palabras, ¿reflexionarán los animales sobre las cosas que pasan en su vida o solo actúan de manera mecánica, respondiendo instintivamente a distintos estímulos?

En *Acerca del alma*, Aristóteles diferenció tres niveles de intelecto: el *alma vegetativa*, que poseen las plantas y el resto de los seres vivos, se ocupa de la procreación y el alimento; el *alma sensitiva* abarca las sensaciones y es característica de los animales porque cualquier animal al menos posee el sentido del tacto, mientras que el *alma intelectiva* permite el razonamiento y es exclusiva de los humanos.

Conforme a esta visión, la capacidad de razonar es lo que nos distingue del resto de los animales y de allí la famosa caracterización del hombre como “un animal racional”.

En línea con los pensamientos de Aristóteles, los filósofos estoicos de la antigua Roma también creían que los animales no están dotados de *logos* (entendimiento) y que se comportan siguiendo un instinto natural de supervivencia. Por lo tanto, los veían como seres creados para el beneficio del hombre y no sentían ningún tipo de obligación moral hacia ellos. Esta concepción de los animales era crucial para los estoicos, ya que, basando su filosofía en el desarrollo de la virtud, si llegaran a dudar de la irracionalidad de los animales tendrían que replantearse el trato que le dispensaban y los fundamentos mismos de su doctrina.

Una crítica visceral a esta postura fue hecha por Plutarco [46-120 d. C.] en sus *Moralia* (*Obras morales y de costumbres*), una recopilación de 78 ensayos en donde discute una amplia variedad de temas, desde la inconveniencia de contraer deudas, hasta las caras visibles de la Luna o la existencia del alma según Platón. En varios de estos ensayos Plutarco busca demostrar la capacidad de raciocinio de los animales, argumentando que estos tienen inteligencia, aunque no tan desarrollada como la de los humanos. Al respecto, dice en “De solertia animalium” (“Sobre la inteligencia de los animales”, *Moralia*, 963_b):

> Pues muchos animales, amigo mío, superan a toda la humanidad por su tamaño o su rapidez, por su vista penetrante o su fino oído, pero no por ello es el hombre ciego ni inválido ni carente de oídos; también corremos, aunque sea más despacio que los ciervos, y vemos, aunque peor que los halcones; la naturaleza no nos ha privado de fuerza ni de volumen, por más que en estos aspectos no somos nada comparados con el elefante y el camello. Así que, si los animales tienen una actividad mental más torpe y discurren peor, no digamos tampoco que no discurren ni tienen actividad mental en absoluto, ni que carecen de racionalidad, sino que poseen una razón débil y turbia, como un ojo con una visión escasa y perturbada.

Plutarco ilustra esta idea describiendo el comportamiento de varios animales. Por ejemplo, relata que los habitantes de la antigua Tracia

utilizaban zorros para verificar la solidez del hielo al cruzar un río helado, pues el zorro avanza despacio y aplica el oído al suelo para estimar el grosor de la capa de hielo a partir de escuchar la corriente que fluye por debajo (*Moralia*, 969a). Por otro lado, afirma que un perro que debe elegir entre tres caminos para seguir a otro animal, al no encontrar rastros en los dos primeros, infiere que el animal ha tomado el tercer camino. Con estos ejemplos, Plutarco busca demostrar que los animales no solo usan sus sentidos, sino también la razón para tomar decisiones.

Las ideas de Plutarco son retomadas por Sexto Empírico (160-210 d. C.) en sus *Esbozos pirrónicos*, un tratado que sigue la tradición del escepticismo de Pirro cuestionando las verdades objetivas, y luego por Michel de Montaigne (1533-1592) en *Apología de Raimundo Sabunde*, en donde el filósofo francés se replantea, también con una gran dosis de escepticismo, los dogmas de la Iglesia Católica. En este ensayo, Montaigne desarrolla una férrea defensa del raciocinio de los animales, dando una infinidad de ejemplos de su inteligencia y célebremente se pregunta: “Cuándo juego con mi gato, ¿cómo sé que soy yo quien juega con él y no él que juega conmigo?” (*Apología de Raimundo Sabunde*, p. 389).

Plutarco (izquierda), Sexto Empírico (centro) y Michel de Montaigne (derecha), tres célebres defensores de la capacidad de inteligencia animal.

Montaigne despreciaba la arrogancia y, en particular, la vanidad

con la que el hombre se ve a sí mismo comparado con otros animales. Era, por ende, muy crítico de la visión antropocéntrica del universo pregonada por el cristianismo, en donde los humanos son considerados la especie elegida que tiene la capacidad única de inteligencia y pensamiento. Sus palabras en *Apología de Raimundo Sabunde* (pp. 386 y 388) no dejan lugar a duda:

> […] sobre qué cimientos ha edificado la superioridad inmensa que cree disfrutar [el hombre] sobre las demás criaturas. ¿Quién le ha enseñado que ese movimiento admirable de la bóveda celeste, el eterno resplandor de esas antorchas que soberbiamente se mantienen sobre su cabeza, las tremendas sacudidas de esa mar infinita, hayan sido establecidos y continúen durante siglos y siglos para su comodidad y servicio? ¿Es acaso posible imaginar nada tan ridículo como esta miserable y raquítica criatura que ni siquiera es dueña de sí misma, que se halla expuesta a recibir daños de todas artes, y que, sin embargo, se cree emperadora y soberana del universo mundo, del que ni siquiera conoce la parte más ínfima, lejos de poder gobernarlo? […]
> ¿Por qué, pues, los consideramos [a los animales] como privados de alma, vida y raciocinio? ¿Acaso hemos podido reconocer en ellos la inmovilidad y la insensibilidad, no habiendo con ellos mantenido otra relación que la de sumisión y obediencia? ¿Osaremos decir acaso que no hemos visto en ninguna criatura si no es en el hombre el empleo de un alma razonable?

Un par de años después de la muerte de Montaigne, nacería el más brillante de los filósofos franceses, René Descartes, quien daría una visión radicalmente distinta sobre la eventual inteligencia de los animales. Este pensador postuló la división entre el cuerpo y la mente; el cuerpo es el dominio de lo físico, mientras que la mente, exclusivamente humana, es el sostén de nuestro pensamiento. Para Descartes los animales son como autómatas que parecen actuar como si fueran conscientes, pero que en realidad no lo son, ya que solo reaccionan mecánicamente basándose en reflejos, sin la capacidad de reflexionar sobre sus actos.

Como Aristóteles, Descartes no negaba que los animales pudieran sentir, sino razonar. Según su visión, el comportamiento de los animales y su capacidad de percepción puede explicarse a partir de principios mecánicos que no requieren entendimiento. En la quinta

parte de su monumental *Discurso del método* fundamenta esta postura en dos hechos. Primero, los animales carecen de un lenguaje articulado, que es el medio para expresar el pensamiento, a partir de lo cual concluye que no pueden mostrar su pensamiento simplemente porque no lo tienen. Segundo, el comportamiento animal es limitado a tareas específicas, mientras que el comportamiento humano está basado en la razón, un instrumento universal que puede usarse en distintas coyunturas. (Nótese la relación con los argumentos de Minsky y la discusión sobre la inteligencia general al final del primer capítulo.)

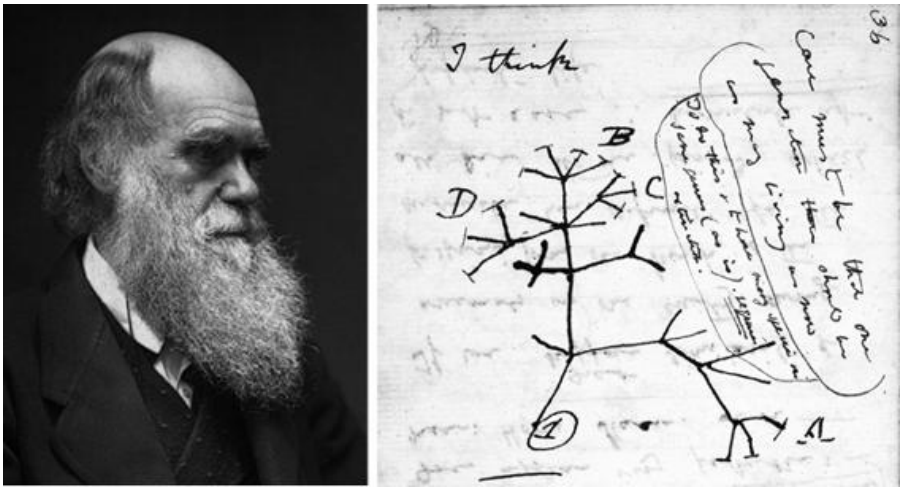
En la cita anterior, Plutarco argumenta que los animales pueden incluso superar a los humanos en diversas tareas. Descartes refuta este argumento alegando que esto no prueba que tengan ingenio, ya que es análogo a decir que un reloj, compuesto de ruedas y resortes, puede contar las horas mejor que nosotros (*Discurso del método*, V). Montaigne, por otro lado, decía que los animales también se comunican entre ellos, aunque no los entendemos, "así como no entendemos a los extranjeros que vienen de otras tierras y que consideramos brutos e ignorantes porque no conocen nuestras costumbres y permanecen mudos al no poder hablar nuestro idioma" (*Apología de Raimundo Sabunde*, pp. 389 y 404). Sin embargo, Descartes aduce que los animales tienen órganos parecidos a los nuestros y que podrían darse a entender[40]. En particular, tanto Plutarco como Montaigne destacan la capacidad de habla de las urracas y de los loros, pero Descartes nota que estos animales solo pueden repetir palabras aunque sin "hablar", es decir, sin llegar a expresar pensamientos[41]. Más aún, Descartes agrega que los hombres que nacen sordos o mudos, careciendo de los órganos que permiten el habla, pueden aun comunicarse haciendo uso de signos.

Ninguno de los argumentos de Plutarco, Montaigne y Descartes ofrece una prueba contundente, ya sea a favor o en contra de la inteligencia de los animales. Sin embargo, a mediados del siglo XIX Charles Darwin concebiría la teoría que revolucionaría nuestra visión de los animales y las raíces mismas de nuestra procedencia. En *Sobre el origen de las especies*, publicado en 1859, Darwin propone que con el

paso del tiempo, de generación en generación se van dando cambios en las especies (que hoy atribuimos a variaciones en el ADN). Tales cambios no son necesariamente beneficiosos, pero, debido a un proceso de selección natural, los organismos competirán por los mismos recursos (agua, comida, espacio) y los más aptos sobrevivirán, mientras que aquellos con variaciones menos adaptadas al ambiente se terminarán extinguiendo. Por supuesto este proceso lleva millones de años, pero el resultado final es que, dadas las mutaciones al azar y el proceso de selección natural, las especies evolucionan con el tiempo.

De esta teoría se desprende que todas las especies tienen ancestros en común; formas más sencillas de vida a partir de las cuales evolucionan como las ramas de un árbol que salen de un mismo tronco. Con mis familiares más cercanos (mis hermanos), tengo un ancestro común inmediato (mis padres); con mis primos el ancestro es más lejano (mis abuelos), y cuanto menos relacionados los miembros de mi familia, más lejano el ancestro en común. De la misma manera, las especies más cercanas, como ser el hombre y los primates superiores, tienen ancestros en común relativamente recientes en la escala de evolución, mientras que para encontrar ancestros comunes con aquellas especies con las que tenemos diferencias mayores (por ejemplo, con las aves o los reptiles) hay que ir más atrás en el tiempo.

La teoría de la evolución de Darwin no hace entonces distinción entre el hombre y el resto de los animales. Todos surgimos a partir de la evolución de distintas ramas del árbol de la vida. La noción de evolución por selección natural, una de las ideas más sencillas y brillantes en la historia de la humanidad, fue un tremendo mazazo a los fundamentos mismos de la concepción antropocéntrica de Descartes. ¿Pues si todos descendemos de los mismos ancestros, por qué seremos solo nosotros los únicos capaces de pensar y ser conscientes de nuestra existencia? O, como diría Montaigne: “Y ese privilegio que el hombre se atribuye en este soberbio edificio de pretender ser único en cuanto a capacidad para reconocer la belleza […] ¿quién le ha otorgado semejante privilegio?” (*Apología de Raimundo Sabunde*, p. 386).



Charles Darwin y su ilustración del árbol de evolución de las especies en un cuaderno de 1837. En el texto de la izquierda, Darwin dice que “en una dada generación debería haber tantas especies como las que hay ahora y que esto a su vez requiere de extinción”.

El ADN es el código genético que se transmite de generación en generación en los seres vivos; una especie de manual de instrucciones que determina el desarrollo de las células en un organismo. El ADN de mis hermanos es muy parecido al mío porque descendemos de los mismos padres y tenemos rasgos en común que heredamos de ellos. Pero por más parecido que sea el ADN entre dos personas, siempre es distinto, y estas diferencias determinan las características específicas que nos individualizan. En promedio, la diferencia de ADN entre dos personas cualesquiera es de alrededor del 0.1%. Esta diferencia hace que algunos seamos más altos que otros, que tengamos ciertos comportamientos innatos, ojos oscuros, a veces claros, el pelo enrulado o una mancha de nacimiento.

Para dar una idea de la diferencia que tenemos con distintas especies, a partir de estudios de restos prehistóricos, hoy sabemos que la diferencia entre nuestro ADN y el del hombre de Neandertal, de quien nos separamos hace medio millón de años en el árbol biológico, es de alrededor del 0.5%. La separación con los chimpancés, bonobos y gorilas, nuestros parientes más cercanos, se produjo hace una decena de millones de años y compartimos casi el 99% del ADN con ellos.

Con los orangutanes la diferencia es un poco mayor (del orden del 3%), con el resto de los monos es aún más grande y así sucesivamente. A luz de estos datos, dada la mínima diferencia entre nuestro código genético y el de los primates superiores (chimpancés, bonobos, gorilas y orangutanes) cabe preguntarse, nuevamente, si nuestros primos en el árbol evolutivo podrían ser como nosotros; si hay alguna diferencia fundamental que nos separe de ellos o si tal diferencia se debe a que han sido relegados dado nuestro dominio en la Tierra. De hecho, ¿podrían en una realidad alternativa llegar incluso a superarnos como en *El planeta de los simios*?

En 1872, poco más de 10 años tras haber presentado su teoría sobre la evolución de las especies, Darwin publicó *La expresión de las emociones en el hombre y en los animales*. En este tratado describe la notable similitud en la manera de expresar emociones (es decir, estados mentales) de distintas especies, reafirmando la noción de afinidad biológica dada por la evolución a partir de antepasados comunes. Tiempo después, en 1925, Wolfgang Kohler (uno de los fundadores de la teoría Gestalt en psicología) describió el comportamiento de un grupo de chimpancés llamados Chica, Grande, Konsul y Sultán, que se valían de apilar cajas y unir palos para alcanzar bananas colgadas fuera del alcance de la mano. Lo interesante del caso es que Kohler afirmaba que estos chimpancés no daban con la solución al azar, sino que lo hacían tras un tiempo de reflexión e inspiración, luego de algunos primeros intentos fallidos. En otras palabras, su comportamiento respondía a un proceso de razonamiento y uso de inteligencia. Las conclusiones de Kohler claramente contradicen las hasta entonces predominantes ideas de los científicos de la escuela del *conductismo*, para quienes los animales solo aprenden a asociar estímulos con respuestas específicas, pero sin entender lo que hacen.

Izquierda: Sultán uniendo dos varas para alcanzar un objeto.
Derecha: Grande alcanzando comida usando cajas apiladas.
Imágenes del libro *The mentality of apes* (Wolfgang Kohler; 1925).

La descripción de Kohler me recuerda una hermosísima historia contada por Plutarco (también discutida por Montaigne). Un barbero de la antigua Roma tenía en su establecimiento un arrendajo (pájaro de la familia de los córvidos) que era un prodigio imitando palabras o distintos sonidos. Un buen día, un grupo de músicos se quedó tocando para placer de los transeúntes frente a la barbería y, a partir de ese momento, el arrendajo no volvió a emitir sonido. Se sospechaba que el sonido de los clarines lo había aturdido y que tal déficit en la audición le había afectado la voz. Pero el motivo era muy distinto. Tras escuchar la música, el arrendajo había quedado sumido en una profunda meditación, preparando su voz para imitar los nuevos instrumentos y, efectivamente, tras un tiempo estalló de nuevo en canto, repitiendo la melodía y el sonido de los clarines y demás instrumentos[42].

Curiosamente, el arrendajo ha sido un animal bastante estudiado en neurociencia debido a su increíble capacidad de memoria. Esta ave guarda comida en el verano (bellotas y semillas) para poder sobrevivir en el invierno, pero también debe ocultar la comida en distintos lugares, previniendo que alguno de los escondites sea descubierto. Lo

sorprendente es que estos pájaros recuerdan la localización de miles de lugares distribuidos en un radio de varios kilómetros alrededor de su nido y hasta son capaces de reconocer bajo la nieve el lugar exacto en donde dejaron una bellota meses atrás[43]. Más aún, los arrendajos suelen robarse comida unos a otros y si notan que otro pájaro los observa al momento de guardar la comida, volverán más tarde a cambiar los escondites para engañar al eventual ladrón. Esto demuestra que tienen lo que se conoce como *teoría de la mente*; que son capaces de ponerse en el lugar del otro. En otras palabras, el arrendajo que guarda comida entiende cómo piensa quien lo está observando y a partir de ponerse en su lugar puede luego engañarlo.

La capacidad de ponerse en el lugar del otro, de entender sus pensamientos y emociones, no es nada trivial y es la base de nuestras interacciones sociales, de nuestra comunicación, empatía, enseñanza y engaño. Es algo que históricamente también se atribuía exclusivamente a los humanos y que los bebés recién adquieren tras un par de años de vida.

En un test clásico de psicología, el niño a ser testeado ve que una persona esconde una golosina en una caja, luego deja la habitación y, en su ausencia, otra persona cambia la golosina de lugar. Al volver la primera persona a la habitación, el chico tiene que indicar dónde cree que buscará la golosina. Recién tras algunos años los chicos tienden a señalar el primer lugar, demostrando entender que lo que cree una persona puede diferir de la realidad.

Uno de los resultados científicos más resonantes del 2016 usó una adaptación de este experimento con videos, mostrando claramente que los chimpancés, bonobos y orangutanes también poseen una teoría de la mente y pueden ponerse en el lugar del otro y entender cómo piensa según sus creencias (incluso si son falsas) para predecir dónde van a buscar un objeto que alguien cambió de lugar sin que lo sepan[44]. El hecho de entender que otro animal tiene creencias falsas es la base del engaño. Así como el arrendajo cambia el lugar en donde esconde sus bellotas, los chimpancés suelen engañarse unos a otros, por ejemplo, desviando la atención de un objeto de interés en presencia de otros animales para después acceder a este sin

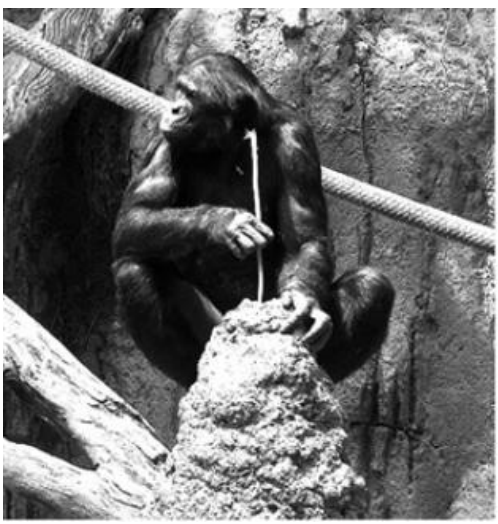
competencia, o distrayendo al macho dominante para cortejar a una hembra[45].

De la misma manera que nos preguntamos si los animales pueden entender el pensamiento de otros, podemos también preguntarnos si tienen consciencia de su propia existencia. Las dos preguntas están de hecho muy relacionadas, porque tendemos a usar el conocimiento que tenemos de nosotros mismos para hacer predicciones de las conductas de los otros. Obviamente no tenemos un lenguaje en común como para preguntar a los animales por su consciencia y eventualmente implementar una evaluación análoga al test de Turing que describiéramos en el capítulo anterior. Sin embargo, el psicólogo norteamericano Gordon Gallup Jr. dio con una solución muy ingeniosa en la década del 70.

Al notar que los chimpancés parecían reconocerse frente al espejo, Gallup diseñó el siguiente test: mientras dormían, les pintó parte de las cejas y las orejas con tintura roja y al despertar los puso frente a un espejo. Gallup notó que los animales comenzaban a tocarse regularmente las partes pintadas, es decir, se reconocían en su imagen reflejada. Por más trivial que parezca, son muy pocos los animales que pasan este test (entre ellos, los primates superiores, los delfines y elefantes)[46]. La mayoría de los animales piensan que su reflejo es otro animal: los pollitos pían constantemente si están solos, pero se calman si están con otros pollitos o frente a un espejo; las gallinas comen más si están con otras gallinas o frente a un espejo; las palomas ponen menos huevos en aislamiento que si están con otras palomas o frente a un espejo, etcétera.

Hemos descripto rasgos indiscutibles de inteligencia en primates superiores, pero tales logros no se limitan a manipular objetos para alcanzar comida (como mostrara Kohler), a usar conductas de engaño o a reconocerse en el espejo. Jane Goodall, quien estudiara a los chimpancés en su hábitat natural en África durante años, observó que también eran capaces de usar herramientas (algo históricamente atribuido exclusivamente a los humanos). Por ejemplo, se valían de ramitas que introducían en un hormiguero para comer termitas, o usaban rocas para romper la cáscara de frutos secos. Tiempo después,

Robert Seyfard y Dorothy Cheney observaron que una especie de monos en su entorno natural emitían distintos gritos de alarma según el peligro. En particular, estos animales tenían un grito específico que usaban al avistar predadores terrestres (un leopardo), otro para predadores en vuelo (un águila) y otro distinto para predadores rastreros (una pitón). Según el grito, las reacciones del resto de los monos eran distintas: trepaban a un árbol ante la alarma de leopardo, miraban hacia arriba y se refugiaban en arbustos al escuchar la alarma de águila, o tomaban una postura bípeda mirando el suelo ante la alarma de pitón[47].



Bonobo extrayendo termitas con una rama en el zoológico de San Diego y mono capuchino usando una roca para romper un fruto seco[48].

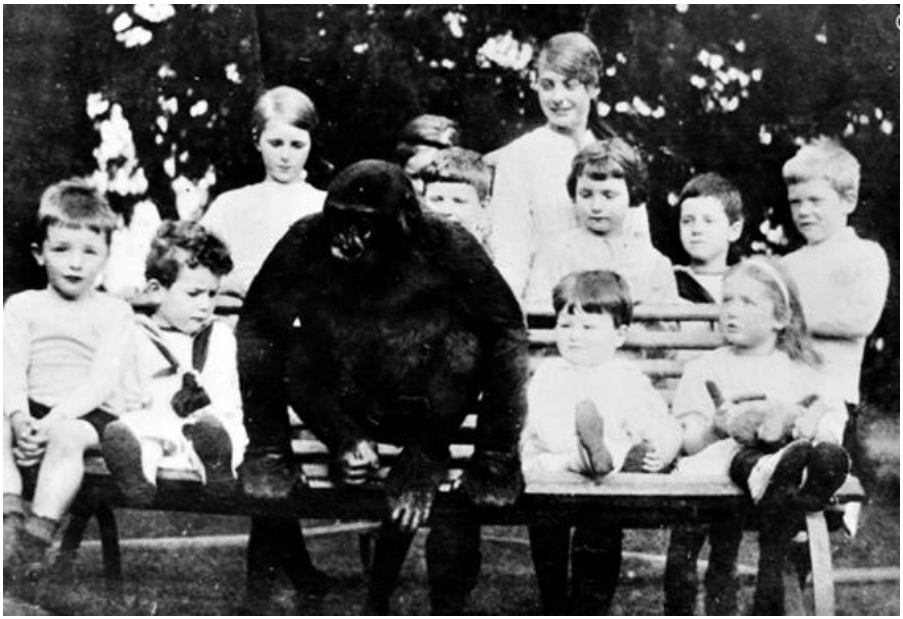
Queda claro, entonces, que la inteligencia no es un atributo único de los humanos y que no existe una diferencia fundamental, sino de grado, con el resto de los animales. Charles Darwin hace una brillante y detallada defensa de este punto en *El origen del hombre* (1871). Dice al respecto:

> No puede caber duda alguna en que existe una diferencia inmensa entre el espíritu del hombre más inferior y del animal más elevado. Si a un mono antropomorfo le fuese posible considerarse a sí mismo de una manera imparcial, podría admitir que, aunque capaz de combinar un plan ingenioso para saquear

> un jardín, o de servirse de piedras para combatir o para romper nueces, estaría fuera del alcance de su inteligencia el pensamiento de trabajar una piedra para convertirla en herramienta. Aún le sería más difícil seguir un razonamiento metafísico, resolver un problema matemático, reflexionar sobre Dios o admirar una imponente escena de la Naturaleza. […]
> Sin embargo, por considerable que sea la diferencia entre el espíritu del hombre y el de los animales más elevados, es solo, ciertamente, una diferencia de grado y no de especie.

Siguiendo esta postura, cabe preguntarse cuánto de la diferencia entre nuestra inteligencia y la de los animales más avanzados es atribuible a una cuestión cultural, de cómo somos educados, y cuánto se debe al proceso de evolución genética[49]. Así como la historia de Tarzán o la de Mowgli, (en *El libro de la selva*) plantea cómo sería la vida de un niño criado fuera de la civilización, podemos análogamente preguntarnos cuán inteligente podría llegar a ser un simio si fuera criado y educado como un humano.

Casualmente, mientras escribo sobre este tema, sale a la luz en los tabloides ingleses la historia de Mr. John Daniel, un peculiar habitante de un pueblito de la campiña inglesa de principios del siglo XX. De chico, John Daniel iba a la escuela y jugaba con sus compañeros, se hacia la cama, ayudaba a lavar los platos y hasta se le permitía tomar el té con los adultos. Pero John Daniel no era un chico más, sino un gorila adoptado a pocos meses de haber nacido, que fue criado como un hijo por una afectuosa señora de buen pasar económico. La increíble historia de este gorila, que disfrutaba particularmente de tomar sidra y era uno más entre los habitantes del pueblo, lamentablemente no tiene un final feliz.



Gorila John Daniel con alumnos de la escuela en un pueblo de Inglaterra, alrededor de la década del 20.(Gloucestershire Live/ SWNS)

Al volverse adulto y aumentar mucho de tamaño y de peso, a esta señora se le hizo imposible seguir manteniéndolo en su casa. Entonces decidió venderlo a una compañía especializada en el comercio de animales, con la promesa de que iría a un parque en Florida. Sin embargo, John Daniel terminó en un circo y murió al poco tiempo de neumonía. Sus cuidadores de aquel entonces atribuyeron su muerte a un estado de depresión que lo llevó a dejar de comer por largos períodos. John Daniel extrañaba su casa y su aldea natal, pasando de haber crecido jugando con el resto de los chicos, a ser una curiosidad expuesta en una jaula.

La historia de John Daniel está muy ligada al argumento de una de las sagas de *El planeta de los simios*. El éxito comercial de la película original dio lugar a cuatro secuelas, estrenadas entre 1970 y 1973, una serie de televisión, distintos cómics, una remake dirigida por Tim Burton en 2001, y una trilogía entre 2011 y 2017. En la primera entrega de esta trilogía, *El planeta de los simios: Revolución* , Will Rodman, el científico de un laboratorio en donde se testean medicamentos para el alzhéimer con monos, descubre a una cría de

pocos días al revisar la jaula de una chimpancé recién sacrificada. Compadeciéndose del pobre animalito, decide llevarlo a su casa a escondidas. El chimpancé, al que llama Cesar, vive en el ático de su casa y tras la paciente instrucción que Rodman le da día a día, con el tiempo comienza a mostrar notables progresos en su aprendizaje.

La idea de criar a un simio como a un humano ha atraído no solo la atención de escritores de ciencia ficción, sino también de muchos investigadores. En la década del 40, Keith and Catherine Hayes, una pareja de científicos que trabajaban en un laboratorio de primates en Florida, notaron que comparar el comportamiento de un humano con el de un simio criado en la selva, o en el confín limitado de un laboratorio, no tenía mucho sentido. Por este motivo, decidieron adoptar una chimpancé llamada Viki a solo días de haber nacido y la criaron como un miembro más de la familia. Los Hayes documentaron en filmaciones los progresos de Viki a medida que iba madurando y observaron que en muchos aspectos la chimpancé se comportaba como un niño de su edad, aprendiendo sobre la base de la imitación.

Viki jugaba como un chico con distintos juguetes y objetos, y con el tiempo aprendió a vestirse sola, ir al baño, comer de un plato usando una cuchara, abrir latas y botellas, servir una taza de café, tomar chocolate en taza y agregarle agua fría si estaba muy caliente, prender fósforos y hasta realizar tareas motoras complejas como usar un martillo, una sierra, una lima de uñas o un destornillador. Replicando lo hecho por los chimpancés estudiados por Kohler, Viki también era capaz de unir varas o apilar cajones para alcanzar objetos colgados de un árbol. Para tener evidencia aún más contundente de su inteligencia, los Hayes trataron de enseñarle a hablar. Sin embargo, en este aspecto su progreso fue mucho más limitado y tras varios años de aprendizaje solo llegó a decir cuatro palabras: “mamá”, “papá”, “up” (arriba) y “cup” (taza).

Tiempo después, en la década del 60, Beatrice y Allen Gardner, una pareja de psicólogos norteamericanos, notaron que la deficiencia en el aprendizaje del habla en los simios se debe a la estructura de su tracto vocal, lo cual no les permite vocalizar como los humanos. Pero los Gardner encontrarían una manera brillante de resolver este problema.

Adoptaron una chimpancé llamada Washoe, a la que, en vez de tratar de hacerla hablar, le enseñaron a comunicarse usando el lenguaje de señas de los sordomudos. Washoe aprendió un par de cientos de signos, los cuales también era capaz de combinar. Por ejemplo, al ver por primera vez un pato señalizó "pájaro de agua" y en otra oportunidad dijo que una sandía era una "fruta golosina".

En las antípodas de lo argumentado con estos experimentos, y en línea con el pensamiento de Descartes, para lingüistas como Noam Chomsky (popularmente conocido tanto por sus posturas científicas como políticas) el lenguaje es una facultad innata y exclusiva del hombre. Steve Pinker (*El instinto del lenguaje*, 1994, p. 18), otro mentado lingüista, ilustra magistralmente este punto con la siguiente analogía:

> […] las personas saben hablar en el mismo sentido en que las arañas saben tejer sus telas. Tejer una tela no es un invento de una araña anónima y genial, ni depende de si la araña ha recibido o no una educación apropiada […]. Las arañas tejen sus telas porque tienen cerebro de araña […]

El documental que muestra las proezas de Washoe (*The first signs of Washoe*) justamente comienza con citas de Chomsky y otros pensadores contemporáneos que los Gardner trataban de refutar. Pero la conclusión de los Gardner no fue universalmente aceptada. Tras analizar en detalle los videos de Washoe, Herbert Terrace, un psicólogo de la Universidad de Columbia, en Nueva York, criticó estos resultados afirmando que el chimpancé solo repetía signos que hacían quienes lo entrenaban, sin demostrar poseer la capacidad de lenguaje. De hecho, el lenguaje va mucho más allá de la atribución de signos a distintos conceptos, pues involucra el uso de la sintaxis para relacionar lógicamente palabras y expresar pensamientos.

Decidido a superar estas críticas y así demostrar que Chomsky y sus colegas estaban equivocados, Terrace planeó un proyecto científico que revelaría la posibilidad de "humanizar" a un mono. La idea era mostrar que un chimpancé educado desde pequeño como un humano podría llegar a hablar usando lenguaje de señas como Washoe. Pero a

diferencia de lo hecho por los Gardner con Washoe, el objetivo primario de Terrace era mostrar que un chimpancé podría entender reglas sintácticas como para formar oraciones simples y diferenciar, por ejemplo, entre "Juan le dio la pelota a Pedro" y "Pedro le dio la pelota a Juan". En el fondo, el fin último era comunicarse con un simio para intercambiar información y vislumbrar qué piensan y sienten los animales más avanzados.

El proyecto terminaría siendo un desastre y no solo desde el punto de vista científico, sino desde el punto de vista ético y humano. La historia del pobre chimpancé, a quien llamaron Nim Chimpsky en clara referencia a Noam Chomsky, es descripta en el documental *Proyecto Nim* (basado en un libro de Elizabeth Hess), que relata la crueldad con la que fue tratado y el desapego y la falta de ética con la que Terrace llevó a cabo su proyecto.

Curiosamente, *Proyecto Nim* y *El planeta de los simios: Revolución* fueron estrenadas al mismo tiempo y sus argumentos tienen notables coincidencias. Pero esas coincidencias no son casuales, pues en un reportaje, uno de los guionistas de *El planeta de los simios* dijo haberse inspirado en la historia de Nim. Imagino, también, que los realizadores del documental quizás hayan planeado estrenarlo al mismo tiempo que la película para aprovechar la inmensa difusión que esta daría a su descripción de la historia real.

Así como en la remake de *El planeta de los simios* Cesar fue adoptado por un científico, Nim fue dado a Stephanie LaFarge, una antigua estudiante de doctorado de Terrace, quien lo crió en su casa junto al resto de sus hijos. Sin embargo, en esa casa nadie dominaba el lenguaje de sordomudos como para poder enseñarle a comunicarse de una manera eficaz. Tras años de vivir como un miembro más de la familia de LaFarge, Terrace decidió mudar a Nim y dejarlo a cargo de otra estudiante. Luego volvió a poner a otra estudiante a su cargo, y así pasaron los primeros años de Nim, viendo cambiar tres veces su tutor y lidiando con distintos estudiantes que trataban de enseñarle el lenguaje de sordomudos en distintas sesiones de entrenamiento.

En la remake de *El planeta de los simios*, Will Rodman, el científico a cargo de Cesar, decide que no tiene más remedio que darlo en

adopción a un albergue de animales tras un altercado con un vecino. La escena en la que Rodman deja a un incrédulo Cesar junto a otros simios es quizás el punto más dramático de la película. (Con el tiempo, Cesar planeará el escape masivo de los simios, a los que llevará a refugiarse en un bosque en las afueras de San Francisco. Estos simios serán la semilla de la nueva civilización que terminará dominando la tierra en el futuro. Pero esa es otra historia.)

En la vida real, Nim no llegaría a organizar ninguna revolución, pero lamentablemente terminaría en un albergue de animales como Cesar. Tras haberse quedado sin financiamiento, Terrace estimó que ya tenía suficientes datos y decidió cancelar imprevistamente el proyecto. Es cierto que Nim se volvió más agresivo de adulto[50] y también es cierto que es difícil continuar un proyecto científico sin financiamiento. Pero esto no es excusa, porque deberían haberse previsto todas estas eventualidades en vez de actuar de una manera tan improvisada e irresponsable. Una de las críticas literarias de la película expone perfectamente el problema de fondo citando un diálogo entre el Principito y el zorro (Antoine de Saint-Exupéry, *El Principito*):

> Los hombres han olvidado esta verdad —dijo el zorro—, pero tú no debes olvidarla. Eres responsable para siempre de lo que has domesticado.

El testimonio de las últimas personas a cargo de Nim, relatando el momento en el que tienen que dejarlo en un albergue de animales, es realmente desgarrador. La tristeza que genera es infinitamente superior a la de la escena análoga en *El planeta de los simios*, porque en este caso la historia es real. Nim sería luego trasladado a un laboratorio en donde lo usarían para testear distintas vacunas y, en cautiverio, trataba infructuosamente de comunicarse con sus cuidadores y los otros chimpancés usando el lenguaje de señas que le habían enseñado.

Mucho tiempo después, Stephanie LaFarge visitó a Nim ya retirado en un santuario de animales. Este la reconoció al instante y sus cuidadores notaron que se puso muy nervioso y potencialmente

agresivo. Desoyendo su opinión, LaFarge insistió con entrar a la jaula. En ese momento, de un dramatismo extremo, encontrándose con su primera madre tras muchos años, Nim al principio se aleja, pero luego la toma del tobillo y comienza a revolearla violentamente por el piso. Sin embargo, no la mata. Según LaFarge, quería demostrarle lo decepcionado que estaba por cómo lo había abandonado a tan miserable e injusto destino.

Personalmente confieso que duele ver *Proyecto Nim* y escribir sobre este tema, pero creo que es absolutamente necesario para recordarnos nuestra obligación, como especie superior, con el resto de los animales. Esto vale particularmente para los científicos. En mi caso, apoyo la investigación con animales y repudio los actos violentos o las opiniones sesgadas de activistas que se oponen dogmáticamente a todo y tratan de infundir temor en la comunidad científica. Estos experimentos son a mi entender necesarios, pero tienen que estar muy cuidadosamente regulados por comités éticos que garanticen que no haya proyectos como el de Terrace. Alguna vez escuché decir al dalái lama (el líder espiritual de los budistas, para quienes todos los animales tienen un cierto grado de consciencia y, consecuentemente, evitan matar hasta un mosquito) que le parecía necesaria la experimentación en animales pero que había que tratarlos con compasión[51]. Sus palabras sintetizan perfectamente mi punto de vista sobre este tema.

Más allá del aspecto ético que acabamos de describir, el proyecto de Terrace fue un rotundo fracaso científico debido a serias falencias. Ante todo, Nim fue entrenado por varias personas distintas, en sesiones de aprendizaje estructuradas, lo cual es análogo a aprender una segunda lengua en el colegio y muy distinto a aprender a hablar a partir de la interacción natural y diaria con una familia de humanos. Por otro lado, el proyecto terminó prematuramente, cuando Nim aún continuaba aprendiendo y, finalmente, la inestabilidad en su ambiente más cercano, cambiando varias veces la persona a cargo de su cuidado, ciertamente no favoreció su aprendizaje. A fin de cuentas, Nim no pudo lograr más que Washoe décadas atrás. Terrace terminó asumiendo su fracaso, dándole la razón al grupo de lingüistas

liderados por Chomsky, quienes afirmaban que los animales no eran capaces de desarrollar el lenguaje y que lo que aparentaba ser comunicación no era más que una imitación de los gestos de sus tutores para hacer distintos pedidos (jugar, comer, tomar agua). Pero la conclusión de Chomsky sería desafiada por un animal extraordinario.

Los bonobos, originarios del Congo, son muy parecidos a los chimpancés, aunque algo más bajos, con un carácter mucho menos agresivo, una postura más erecta y un andar generalmente bípedo. La contextura de un bonobo es, de hecho, muy parecida a la de Lucy, el esqueleto de un australopiteco que fue hallado en Etiopía y que data de hace 3 millones y medio de años. Kanzi, un bonobo nacido en cautiverio, solía jugar y hacer travesuras mientras Sue Savage-Rumbaugh (uno de los pocos científicos que continuó estudiando la capacidad de lenguaje en primates tras el fracaso de Terrace) trataba infructuosamente de enseñarle a Matata, su madre adoptiva, a comunicarse usando un tablero electrónico con distintos símbolos. Pero un buen día, Matata fue temporariamente trasladada a otro centro y Kanzi sintió la irrefrenable necesidad de usar el tablero para preguntar por su madre. Lo sorprendente es que Kanzi nunca había tenido una sesión de entrenamiento y solo aprendió a usar el tablero viendo cómo Savage-Rumbaugh le enseñaba a su madre. La gran diferencia con Nim o Washoe es que Kanzi no usaba signos para pedir algo específico, sino que necesitaba comunicarse. A partir de este episodio, Savage-Rumbaugh entendió que más que enseñarle palabras, debía estimular las ganas de “hablar” del bonobo, compartiendo juegos, comidas, largas excursiones y campings, entre otras muchas cosas.

Con el tiempo, Kanzi aprendió cientos de palabras, comenzó a combinarlas y a entender el significado de oraciones complejas, incluso referidas a hechos del pasado o involucrando objetos y lugares que no estaban a la vista. Justamente una de las grandes limitaciones de no tener lenguaje es el hecho de solo poder intercambiar información sobre el presente inmediato y no sobre experiencias pasadas o planes futuros. Más aún, Kanzi demostró entender el sentido

sintáctico de oraciones, es decir reconociendo que el lenguaje va más allá de una mera asociación de palabras, y que su orden y estructura pueden representar distintos significados. Por ejemplo, cuando en un video Savage-Rumbaugh pidió a otro bonobo llamado Tamule que abrace a Kanzi, Kanzi no abrazó a Tamule sino que tomó su brazo y lo puso alrededor suyo. O sea, Kanzi reconocía la diferencia entre "Tamule, abraza a Kanzi" y "Kanzi, abraza a Tamule". Entre otras muchas cosas, Kanzi también podía cortar ramas y prender el fuego en una excursión, y hasta jugar al Pacman. Savage-Rumbaugh dice que Kanzi también disfruta de ver películas y que justamente una de sus favoritas es *El planeta de los simios*.

Panashiba, hija natural de Matata y media hermana de Kanzi, también demostró poder comunicarse y su fama la llevó incluso a improvisar en el teclado con Peter Gabriel (músico de rock que fuera vocalista de la banda Genesis). Pero la capacidad del lenguaje no se restringe a los bonobos y fue asimismo mostrada por otros primates superiores, como el orangután Chantek o la gorila Koko, en quien Michael Crichton se inspirara al escribir la novela de ciencia ficción *Congo* (luego adaptada a una película con el mismo nombre).

En estas últimas páginas nos enfocamos casi exclusivamente en el lenguaje, ya que nuestro uso del lenguaje aparenta ser la gran diferencia con otras especies. Sin embargo, podríamos argumentar que quizás sea un error enfocarse en el lenguaje como signo casi exclusivo de inteligencia, porque al testear la capacidad de lenguaje en distintos animales puede que estemos cometiendo el error de antropomorfizar, de evaluarlos usando nuestros estándares. Sexto Empírico (*Esbozos pirrónicos*, XIV, 55, 58) afirmaba que los animales tienen distintas representaciones mentales debido a grandes diferencias en sus sentidos.

> Un ungüento aparece calmo muy agradable para los hombres, pero insoportable para los escarabajos y las abejas. El aceite de oliva aprovecha al hombre, pero destruye a las avispas y a las abejas si se vierte sobre ellas. […] Si las mismas cosas para unos son desagradables y para otros placenteras y si lo placentero o desagradable reside en la imaginación, entonces en los animales se dan distintas representaciones mentales de los objetos.

Al igual que Descartes, tendemos a asumir que, como nosotros, cualquier animal inteligente querría expresar sus pensamientos. Pero el hecho de que los animales no hablen entre ellos puede deberse a que tienen distintas representaciones mentales y, al sentir el mundo de otra manera, quizás no tengan la necesidad de comunicarse o tal vez lo hagan por otros medios. El filósofo austríaco Ludwig Wittgenstein (de quien hablaremos en el siguiente capítulo) decía al respecto que "si el león pudiera hablar, no podríamos entenderle". Más allá de estas salvedades, el uso del lenguaje conlleva dos grandes beneficios: por un lado, nos permite comunicarnos y trasmitir conocimiento y, por el otro, refuerza la capacidad de abstracción, de categorizar y formar conceptos. El lenguaje hace que podamos ir más allá de pensar exclusivamente sobre las cosas a la vista en nuestro presente inmediato; nos permite etiquetar conceptos y pensar sobre nuestro pensamiento, evaluar el futuro y meditar sobre el pasado[52]. Más específicamente, las palabras son abstracciones de la realidad. Cuando digo "perro" no me refiero al perro de mi infancia, ni al de mi vecino; no importa si es grande, chico, guardián, cazador o si tiene pelaje claro con manchas en el torso. Al decir "perro" dejo de lado todos estos detalles y me refiero a aquello que define al animal. En un pasaje genial, dice Jorge Luis Borges (*Palabrería para versos*, "El tamaño de mi esperanza", 1926):

> El mundo aparencial es un tropel de sensaciones barajadas… El lenguaje es un ordenamiento eficaz de esa enigmática abundancia del mundo. Dicho sea con otras palabras: los sustantivos se los inventamos a la realidad. Palpamos un redondel, vemos un montoncito de luz color de madrugada, un cosquilleo nos alegra la boca, y mentimos que esas tres cosas heterogéneas son una sola y que se llama naranja. La luna misma es una ficción. Fuera de conveniencias astronómicas que no deben atarearnos aquí, no hay semejanza alguna entre el redondel amarillo que ahora está alzándose con claridad sobre el paredón de Recoleta, y la tajadita rosada que vi en el cielo de la Plaza de Mayo, hace muchas noches. Todo sustantivo es una abreviatura. En lugar de contar frío, filoso, hiriente, inquebrantable, brillador, puntiagudo, enunciamos puñal; en sustitución de alejamiento del sol y profesión de sombra, decimos atardecer.

Más aún, en "Tlön, Uqbar, Orbis Tertius" (*Ficciones*, 1944), Borges concibe un mundo en cuyo lenguaje no hay sustantivos y concluye que en este lugar "Todo estado mental es irreductible: el mero hecho de nombrarlo —*id est*, de clasificarlo— importa un falseo. De ello cabría deducir que no hay ciencia en Tlön —ni siquiera razonamientos"[53].

A fines del siglo XVII, John Locke, uno de los más notables exponentes del empirismo británico, célebremente diría *brutes abstract not* (las bestias no abstraen; *An essay concerning human understanding* II, XI, 10). Locke basaba esta afirmación en el hecho de que los animales no utilizan palabras o algún tipo de signos para representar ideas generales. Hasta cierto punto, hoy día podríamos refutar la afirmación de Locke con casos como los de Kanzi, Koko o Chantek. Los animales y en particular los primates superiores son mucho más inteligentes de lo que creyeran Aristóteles y Descartes, y tienen al menos una limitada capacidad de lenguaje. Sin embargo, ninguno de estos animales escribe poesía, se pregunta por el origen del Universo o resuelve integrales. Esto nos lleva a plantearnos si la diferencia entre humanos y simios es solo una cuestión de medio ambiente y uso de lenguaje, o si existen diferencias más fundamentales.

En la primavera de 1828, un adolescente fue encontrado deambulando perdido en las calles de Nuremberg, Alemania. El chico, de unos 16 años, apenas hablaba o tenía conocimiento de sí mismo. Decía llamarse Kaspar Hauser y aseguraba haber vivido toda su vida confinado en una celda oscura, sin contacto alguno con otra persona. La historia de Kaspar fue una usina de rumores y tuvo en vilo a toda Alemania durante mucho tiempo. Llegó incluso a creerse que era el príncipe heredero de Baden, algo que mucho tiempo después fue desmentido gracias a un análisis de ADN.

La verdadera historia continúa siendo un misterio digno de un cuento de los hermanos Grimm. Hay también quienes dicen que Edgar Rice Burroughs se basó en una historia real al escribir *Tarzán*, y existen asimismo muchos otros relatos de hombres que crecieron fuera de la civilización, desde la leyenda de Rómulo y Remo (los fundadores de Roma) criados por una loba, hasta el caso más reciente de un vietnamita a quien su padre llevó recién nacido a vivir en la selva

escapando de la guerra. Más allá de cuánto queramos creer o no estas historias, es evidente que un humano criado en soledad, completamente alejado de la civilización, no podría alcanzar el desarrollo intelectual de una persona normal. Sin embargo, tal persona podría rápidamente aprender una infinidad de tareas que serían imposibles para un bonobo como Kanzi, aun habiendo crecido rodeado de humanos.

Retrato de Kaspar Hauser.

No niego en absoluto la inteligencia de los animales, pero claramente hay una gran diferencia con las capacidades del hombre.

Esta diferencia no es solo de educación sino también genética. Probablemente haya surgido a partir de muchísimos de años de evolución en donde una única especie, los humanos, hemos ido desarrollando una capacidad de pensamiento más profundo y abstracto facilitado por uso del lenguaje.

En el capítulo anterior describí el descubrimiento de las neuronas de concepto en un área que es clave para la memoria en el cerebro humano. Alguna de estas neuronas respondió a Jennifer Aniston, otra disparó al mostrarle distintas fotos de Halle Berry y otra al presentarle a Oprah Winfrey. Lo importante es que estas neuronas respondieron selectivamente a fotos de personas específicas e incluso a sus nombres dichos o escritos; en otras palabras, no codifican detalles sino conceptos. Y tiene sentido que haya neuronas representando conceptos en un área ligada a la memoria porque tendemos a quedarnos con el significado de las cosas y a olvidar detalles. Esta era la diferencia que remarcáramos con las computadoras y que creo, también, es una de las grandes diferencias con otros animales. Lo interesante es que todavía nadie encontró neuronas como estas en otras especies. Puede que en unos años alguien reporte una codificación análoga de conceptos en monos, pero hasta ahora, y tras varios intentos en distintos laboratorios, no ha sido el caso.

¿Pero tenemos lenguaje porque codificamos conceptos o codificamos conceptos porque tenemos lenguaje? Me atrevo a especular que ambas cosas evolucionaron en paralelo, reforzándose la una a la otra. Aclaro que este argumento es muy especulativo; no es más que una opinión personal, aunque me resisto a pensar que el hecho de que seamos la única especie con lenguaje y con neuronas de concepto (al menos por ahora) se deba a una coincidencia.

Quizás lo más parecido a las neuronas de concepto sean las *place cells* (neuronas que representan lugares específicos) en roedores. Sin embargo, al hacer cambios en el ambiente en donde el roedor realiza un experimento, estas neuronas muestran lo que se conoce como *remapping* (remapeo), es decir, comienzan a responder en lugares distintos y, típicamente, al cambiar el ambiente, el roedor que aprendió a realizar una tarea compleja vuelve a aprenderla

prácticamente de cero[54]. En otras palabras, los roedores no son capaces de generalizar y darse cuenta de que están haciendo exactamente lo mismo, pero en una situación algo distinta.

En general, los roedores tienden a aprender las tareas independientemente, sin buscar relación entre ellas o atajos para facilitar su aprendizaje. La capacidad de generalizar, de aplicar el conocimiento adquirido en una situación a circunstancias distintas, es justamente en lo que prevalecemos los humanos. Como vimos en los capítulos anteriores, una diferencia análoga existe entre el desarrollo de máquinas y algoritmos para tareas específicas, y lo esquivo que hasta ahora resulta el poder crear circuitos capaces de abstraer, generalizar y demostrar inteligencia general. Pero tal inteligencia no es gratis. De todas las especies, somos la que más tiempo tarda en madurar. La diferencia es que las crías de animales comienzan desde su nacimiento a aprender tareas específicas para sobrevivir. Un bebé, en cambio, tiene un desarrollo muchísimo más lento porque está justamente estableciendo las bases de su inteligencia. A corto plazo no podrá lidiar por sí mismo con las necesidades más elementales, pero a largo plazo terminará hasta preguntándose por su lugar en el universo y qué lo diferencia de otras especies.

40 Veremos que esto en realidad no es cierto. En particular, la disposición del tracto vocal de los monos no les permite vocalizar como los humanos.

41 Hoy día sabemos que este argumento de Descartes tampoco es correcto. En particular, un loro llamado Alex, estudiado por la científica Irene Pepperberg, evidencia comprender lo que dice. Por ejemplo, cuando Pepperberg le pregunta de qué color es un determinado objeto, Alex no repite mecánicamente la pregunta, como otros loros, sino que da la respuesta adecuada. Según Pepperberg, entre otras cosas Alex puede reconocer unos 50 objetos, distintas formas y colores, conceptos como “mayor”, “menor”, “igual” y “distinto”, y cantidades hasta 6.

42 Plutarco, *Moralia* $973_{c\text{-}e}$. Montaigne, *Apología de Raimundo Sabunde*, 402.

43 Para una discusión a nivel divulgativo de los experimentos que se hacen con arrendajos, véase:

- V. Morell. Nicky and the jays. *Science* 315: 1074-1075; 2007.

Y para una discusión más detallada y técnica, véase:

- U. Grodzinski y N. Clayton. “Problems faced by food-caching corvids and the evolution of cognitive solutions”. *Philosophical Transactions of the Royal Society*, Series B: Biological Series 365: 977-987; 2010.

44 Los resultados de este trabajo fueron publicados en: Christopher Krupenye, Fumihiro Kano, Satoshi Hirata, Josep Call and Michael Tomasello. Great apes anticipate that other individuals will act according to false beliefs. *Science* 354: 110-114; 2016.

45 Para un análisis detallado de distintas conductas de engaño en primates superiores ver: Richard Byrne and Andrew Whiten. Cognitive evolution in primates: evidence from tactical deception. *Man* 27: 609-627; 1992.

La idea de ponerse en el lugar del otro para poder engañarlo es la quintaesencia de la magia. El ardid de los magos justamente consiste en implantar falsas creencias en el espectador que dan lugar a resoluciones inesperadas. Para más detalles, ver por ejemplo: Rodrigo Quian Quiroga. “Magic and cognitive neuroscience”. *Current Biology* 26: R387-R407; 2016.

46 Para más detalles, ver por ejemplo: Gordon Gallup Jr. “Chimpanzees: Self-recognition”. *Science* 167: 86-87; 1970.

47 Para más detalles de estos y otros experimentos con monos, ver por ejemplo el capítulo IV de *La conciencia viviente* (José Luis Díaz, Fondo de Cultura Económica, 2007), o el capítulo V de *Los dragones del Edén* (Carl Sagan, Crítica, 2002).

48 Imagen del mono capuchino reproducida de: Fragaszy *et al.* “Wild Bearded Capuchin Monkeys (*Sapajus libidinosus*) Strategically Place Nuts in a Stable Position during Nut-Cracking”. *PLOS* One 8: e56182; 2013.

49 Esto es parte de una discusión mucho más general y renombrada que en inglés se conoce como *nature versus nurture* (naturaleza versus educación).

50 Como describe Jane Goodall, la gran dificultad de criar chimpancés en hogares humanos es que estos tienden a volverse agresivos al madurar.

51 Recuerdo escuchar al dalái lama decir esto en respuesta a la pregunta de un miembro de la audiencia, tras una charla que diera en el encuentro anual de la encuentro anual de la Society for Neuroscience en 2005.

52 El filósofo Daniel Dennett hace una brillante descripción de estas ideas en el capítulo 5 de su libro *Kinds of minds* (Basic Books, 1996).

53 Para una descripción más comprehensiva sobre los pensamientos de Borges y otros filósofos sobre la abstracción y el lenguaje ver los capítulos 2 y 12 de *Borges y la memoria*, op. cit.

54 Ver, por ejemplo: Pavel Itskov, Ekaterina Vinnik and Mathew Diamond. “Hippocampal representation of touch-guided behavior in rats: Persistent and independent traces of stimulus and reward location”. *PLOS* One 6 (1): e16462; 2011.

IV. Matrix

Ilusión de la realidad

Neo se ha quedado dormido, escuchando música con los auriculares puestos frente a la computadora. Su escritorio rebalsa de papeles, teclados, cables, impresoras, discos con datos y demás *gadgets*, y el monitor muestra secuencias de búsquedas *online* automáticas, propias de un hacker. De repente la pantalla queda en blanco. Neo despierta y ve aparecer, línea por línea, un críptico mensaje en letras verdes fluorescentes, parecidas a las de los primeros monitores de computadora:

> [Wake up, Neo…
> The Matrix has you…
> Follow the white rabbit.]

> Despierta, Neo…
> Matrix te posee…
> Sigue al conejo blanco.

Semidormido, trata con distintos comandos de quitar el mensaje de la pantalla y aparece nueva línea —knock, knock, Neo— justo en el momento en que golpean la puerta. Alguien viene buscar un disco con datos evidentemente pirateados[55]. Entre el grupo de jóvenes que lo acompañan, una chica con un conejo blanco tatuado en su hombro lo invita a salir con ellos. Neo, intrigado, decide acompañarlos a una discoteca. Allí conocerá a Trinity, la hacker que escribió el mensaje en su computadora, quien eventualmente lo llevará a encontrarse con Morpheus. A Neo lo obsesiona una pregunta —¿qué es la Matrix?— y sabe que Morpheus tiene la respuesta.

En la habitación de un hotel abandonado se encuentra con Morpheus. Sentados frente a frente en dos sillones de cuero, Morpheus comienza a hablarle de la Matrix, pero sus afirmaciones son muy

vagas: “La Matrix está en todos lados, alrededor nuestro, incluso en esta habitación… es el mundo que ha sido puesto frente a tus ojos para ocultarte la realidad… que eres un esclavo. Como todos los demás naciste en cautiverio, dentro de una prisión que no puedes oler, gustar o tocar. Una prisión de tu mente”. Finalmente, Morpheus abre la mano y le muestra dos píldoras. Tomando la píldora azul —dice—, la historia termina y despertará en su cama pensando que todo fue un sueño. Tomando la píldora roja, en cambio, accederá a la verdad y verá por sí mismo qué es la Matrix.

Como era de esperarse, Neo toma la píldora roja y a partir de ese momento es rescatado de la ilusión que hasta entonces era su vida, para enfrentar una realidad desoladora. Cuando se recupera de la atrofia de sus músculos, tras nunca haber usado su cuerpo, finalmente se entera de la verdad. La Matrix es una construcción creada por una Inteligencia Artificial destinada a aplacar la mente de los hombres, a quienes mantiene desde su nacimiento encerrados en cubículos para así poder generar energía a partir de sus impulsos nerviosos. Morpheus le muestra a Neo cómo funciona la Matrix, usando distintas simulaciones de la realidad implantadas directamente en su cerebro. Al experimentar dichas simulaciones, Neo alega que eso no es real. Morpheus toma el control remoto y mostrándole en una pantalla de televisión distintas imágenes de la realidad que conoce, le responde:

> [What is real? How do you define real? If you’re talking about what you feel, taste, smell, or see, then real is simply electrical signals interpreted by your brain…
> This is the world you know. The world as it was at the end of the twentieth century. It exists now only as part of a neural-interactive simulation that we call the Matrix.]

> ¿Qué es “real”? ¿Cómo lo definirías? Si te refieres a lo que puedes sentir, degustar, olfatear o ver, entonces lo real es simplemente señales eléctricas interpretadas por tu cerebro…
> Este es el mundo que conoces. El mundo como era a finales del siglo XX. Existe ahora solo como parte de una simulación neuro-interactiva que llamamos Matrix.

La idea de *Matrix* no es del todo original. De hecho, la noción de una realidad alternativa en el ciberespacio es la base de obras como *Neuromancer*, de William Gibson, uno de los grandes pilares del movimiento *cyberpunk*. Esta corriente dentro de la ciencia ficción plantea el contraste entre los inmensos avances de la tecnología (y el uso del ciberespacio), con la pobre realidad de individuos marginales, típicamente hackers, que desafían el sistema y las grandes corporaciones. El género de cuentos y películas *cyberpunk* nace como una reacción a las obras clásicas de ciencia ficción, describiendo epopeyas de viajes interestelares y futuras civilizaciones. Las obras *cyberpunk*, en cambio, tienden a reflexionar sobre el impacto de la tecnología en lo que significa ser humano.

Más allá de su falta de originalidad, *Matrix* también fue criticada por su exceso de adrenalina, tal vez abusando de escenas con combates de kung-fu o enfrentamientos con armas automáticas que nunca dejan de disparar.

En neurociencia existe un tipo de memoria llamada *flashbulb*. Son los recuerdos de sucesos con un altísimo contenido emocional, ya sea positivo o negativo, que podemos recrear con una increíble riqueza de detalle años después de haberlos vivido. Una memoria *flashbulb* es, por ejemplo, el gol de Maradona a los ingleses en el 86, o la final contra Holanda en el 78. Más de una vez he argumentado que, al menos para el argentino típico, la gran mayoría de los recuerdos *flashbulb* están ligados a partidos de fútbol. Para ilustrar entonces cuánto me gustó *Matrix*, basta decir que recuerdo exactamente el cine donde vi la película hace más de 15 años y cómo terminé deambulando por las calles de Aachen (en esa época estaba haciendo un posdoctorado en Alemania) enloquecido con la historia que acababa de ver y la música final de Rage Against the Machine, repitiéndose una y otra vez en mi cabeza. Poco me importó la eventual falta de originalidad o el exceso de acción. Disfruté como loco.

Pero la idea subyacente en *Matrix* no solo se remonta a las raíces del movimiento *cyberpunk*, sino que va mucho más atrás en el tiempo. Justamente, una de las preguntas más trascendentales de la filosofía

tiene que ver con la existencia del ser y del mundo exterior. Es tan simple y a la vez tan profundo como preguntarse: ¿existo?, ¿existe lo que veo a mi alrededor?

En el libro VII de *La República*, Platón plantea su famosa alegoría de la caverna: un grupo de personas viven encadenadas en una caverna y solo pueden ver las sombras de lo que acontece detrás de ellas; para esta gente dichas sombras son su única realidad, ya que no pueden vislumbrar aquello que las genera. De esta manera, Platón argumenta que solo vemos manifestaciones de la realidad, pues la realidad en sí misma está más allá de nuestra percepción. Por ejemplo, si vemos uno u otro caballo, percibimos manifestaciones específicas a partir de las cuales abstraemos el concepto universal "caballo", que posee una existencia independiente de las distintas manifestaciones e incluso de nosotros mismos.

Detalle de grabado sobre la alegoría de la caverna (Jan Saenredam; 1604). Las personas de la caverna (en la parte inferior) solo pueden ver las sombras proyectadas en la pared (derecha) y no aquello que las genera (izquierda).

Esta concepción filosófica, conocida como "realismo platónico", afirma la existencia de una realidad universal. Para Platón, la tarea del

filósofo es hacer uso de su razón para acceder a la realidad exterior (el concepto de "caballo") a partir de las distintas manifestaciones que le llegan a través de los sentidos (los distintos caballos que observa).

Así como Platón comparara las impresiones sensoriales con las sombras que ven los hombres encadenados en la caverna, dos milenios más tarde Descartes llevaría al límite la idea de dudar de la realidad. En su cruzada por encontrar las verdades últimas, aquellas que no pueden ser refutadas, Descartes dudó sistemáticamente de todo. A tal fin, en la primera de sus *Meditaciones* concibió la posibilidad de un genio maligno y poderoso que utiliza todos sus recursos para engañarnos acerca del mundo exterior. Cabe aclarar que, aunque consideremos esto muy poco probable, es un argumento filosófico ingenioso y válido, porque en principio no podemos refutar la posibilidad de que "el cielo, el aire, la tierra, las figuras, los colores, los sonidos y todas las demás cosas exteriores no son más que ilusiones y sueños de los que se ha servido [el genio maligno] para atrapar nuestra credulidad" (Descartes, primera Meditación, 10). Sin embargo, Descartes argumentó que, por más sofisticado que sea el engaño, podemos aún confirmar nuestra propia existencia ya que somos capaces de plantearnos la duda. Es entonces nuestra capacidad de pensamiento, según Descartes, la que nos asegura que existimos y a partir de esta idea enuncia la frase más mentada de la historia de la filosofía: *Cogito ergo sum* (Pienso, luego existo).

Con los argumentos anteriores, Descartes demuestra la existencia de sí mismo. En las siguientes *Meditaciones* procederá también a demostrar la existencia del mundo exterior, argumentando que Dios no es falaz y que no tiende a engañarnos porque eso lo haría imperfecto. La existencia tanto del pensamiento como del mundo exterior, eventualmente lo llevaría a separar entre mente y materia, lo cual, como vimos en el primer capítulo, es el principio fundamental del dualismo cartesiano.

La postura opuesta al dualismo es el monismo —o sea, afirmar que solo existe el reino de lo material (materialismo) o el de las ideas (idealismo ontológico). En la antigua Grecia, Demócrito (460-370 a. C.), precursor de la teoría atómica y uno de los primeros exponentes

del materialismo, afirmaba que no hay más que átomos y el espacio entre ellos. La visión moderna del materialismo es asumir que la mente no es más que la actividad de neuronas, como ya discutiéramos en detalle al hablar sobre HAL 9000 y la capacidad de pensamiento de las máquinas. El idealismo ontológico afirma exactamente lo contrario: lo único que existe son las ideas y la mente. Su impulsor más ferviente fue el obispo George Berkeley (1685-1753), uno de los exponentes más notables (junto a John Locke y David Hume) del empirismo británico. Para estos filósofos, la mente es como una *tabula rasa* (una pizarra en blanco), en la cual vamos grabando conocimiento a través de la experiencia y la percepción con los sentidos.

Así como el principio rector de Descartes era *cogito ergo sum*, el de Berkeley sería *esse est percipi* (ser es ser percibido). Berkeley argumentaba que la realidad exterior no es más que un conjunto de ideas y sostenía que solo existe nuestra percepción de las cosas y no las cosas en sí mismas. Para Berkeley, aquello que no es percibido directamente no existe. Pero, por supuesto, también existen las mentes que perciben y piensan[56].

Más tarde, Immanuel Kant (1724-1804) propondría una posición menos radical que hoy conocemos como idealismo trascendental. Kant argumentó que solo llegamos a conocer nuestras representaciones de las cosas, pero nunca *die Dinge an sich* (la cosa en sí misma). Tales representaciones se generan a partir de una intuición innata del espacio y del tiempo, es decir, concebimos las cosas en un tiempo y un lugar. Sin embargo, a diferencia de Berkeley, Kant no niega la existencia de las cosas (y por ende, su postura no es opuesta al dualismo); solo argumenta que no podemos llegar a conocerlas en sí mismas, independientemente de la percepción que tenemos de ellas[57].

Descartes usa el argumento del genio maligno para poner a prueba (y luego demostrar) su propia existencia, y a partir de creer en un Dios benevolente que no busca engañarnos, también reafirma la existencia del mundo exterior. Sin embargo, si dejamos de lado las conclusiones basadas en la existencia de un Dios benevolente, el argumento del genio maligno nos lleva de narices a Matrix o lo que en filosofía se conoce como solipsismo[58]. Aunque cabe aclarar una diferencia.

Mientras el argumento de *Matrix* estipula que infinidad de personas viven una realidad ficticia implantada en sus mentes, el solipsismo es el idealismo llevado al extremo: no solo no existe el mundo exterior, sino que tampoco existe otra mente más que la nuestra. Parece descabellado, pero ¿cómo refutarlo?, ¿cómo probar que realmente existen las personas que veo a mi alrededor?

No debería sorprendernos que no exista registro alguno de un filósofo solipsista. Más allá de las falencias que describiremos más adelante, pregonar el solipsismo sería una contradicción. Si una persona realmente creyera ser la única en el universo, ¿qué sentido tendría diseminar sus ideas y tratar de convencer de algo a quienes no son más que una ilusión de sus sentidos? Dice Bertrand Russell (*Human Knowledge: Its Scope and Limits*, 1948):

> [As against solipsism it is to be said, in the first place, that it is psychologically impossible to believe, and is rejected in fact even by those who mean to accept it. I once received a letter from an eminent logician, Mrs. Christine Ladd-Franklin, saying that she was a solipsist, and was surprised that there were no others. Coming from a logician and a solipsist, her surprise surprised me.]

> En contra del solipsismo debe decirse que, en primer lugar, es psicológicamente imposible de sostener, y es de hecho rechazado aun por aquellos que aducen aceptarlo. Una vez recibí una carta de una respetada experta en lógica, Christine Ladd-Franklin, diciendo que era solipsista, y que se sorprendía que no hubiera otros. Viniendo de una experta en lógica y una solipsista, me sorprendió su sorpresa.

La versión moderna del genio maligno de Descartes y el solipsismo es lo que se conoce como el argumento de *brain in a vat* (cerebro en un frasco), pergeñado por el filósofo norteamericano Hilary Putnam en 1981[59]. Imaginemos que un científico malvado (o el genio de Descartes) nos somete a una operación con el fin colocar nuestro cerebro en un gran frasco conteniendo una solución saturada de nutrientes para mantenerlo vivo. Imaginemos también que conecta todas nuestras terminales nerviosas a una poderosísima supercomputadora capaz de simular la realidad exterior, a partir de

complejas estimulaciones eléctricas de distintos grupos de neuronas. Por ejemplo, la supercomputadora estimula neuronas que representan los conceptos de arena, palmeras y mar, y nos lleva a creer que estamos en una playa. El argumento, parecido al de *Matrix*, es algo alocado pero muy interesante, pues si carecemos de argumentos para demostrar que no somos un cerebro en un frasco, entonces no podemos afirmar nada sobre la existencia del mundo exterior.

Ilustración del argumento de *brain in a vat*. A través de estímulos provistos por una supercomputadora, el cerebro cree, por ejemplo, estar en una playa.

En neurociencia existen varios ejemplos de cómo la estimulación directa del cerebro puede alterar e incluso generar distintas sensaciones. Por ejemplo, al estimular eléctricamente en monos un

área ligada a la percepción del movimiento (que en neurociencia se conoce como MT o V5), científicos de la Universidad de Stanford mostraron que tal estimulación cambiaba la percepción de los animales sobre la dirección de movimiento de un grupo de puntos en la pantalla[60]. Asimismo, un grupo de investigadores iraníes mostró que la estimulación eléctrica de un área dedicada al reconocimiento de caras (corteza inferotemporal) hacía que los monos tendieran a reportar que las imágenes borrosas que se les mostraba correspondían a caras y no a otros objetos[61]. Pero los estudios más espectaculares describiendo los efectos de la estimulación eléctrica del cerebro fueron hechos hace más de medio siglo por el neurocirujano Wilder Penfield en humanos[62].

En el capítulo 2 mencionamos que los pacientes epilépticos que no responden a medicación pueden ser sometidos a una cirugía para remover el área que provoca las crisis epilépticas. El éxito de tales cirugías depende, por un lado, de localizar con la mayor precisión posible el foco epiléptico (y como viéramos antes, para ello se procede en muchos casos a implantar electrodos intracraneales) y por el otro, de no dañar áreas que son vitales para el funcionamiento del paciente. Por ejemplo, si por error se comprometiera parte de la corteza motora en la resección quirúrgica, el paciente sufriría una parálisis; si se comprometiera la corteza visual, podría quedar ciego, etcétera. Debido a que la ubicación de estas áreas vitales varía de un paciente a otro, Penfield desarrolló un protocolo de estimulación eléctrica para delimitar su localización exacta. Este procedimiento era realizado con el paciente despierto y su cerebro al descubierto (luego de haber removido temporariamente la parte superior del cráneo), lo cual se llevaba a cabo con anestesia local ya que el cerebro mismo no posee receptores de dolor. Penfield estimulaba entonces alrededor de la corteza motora y sensorial, delineando hasta qué punto la estimulación en estas áreas generaba cosquilleos o movimientos de distintas partes del cuerpo, como los dedos, la lengua o los labios. Análogamente, estimulaba alrededor de áreas dedicadas al lenguaje mientras le pedía al paciente que nombrara diferentes objetos, o estimulaba áreas de la corteza visual y producía distintas

alucinaciones (los pacientes referían ver estrellas de colores, etcétera).

Pero los resultados más espectaculares se produjeron cuando Penfield estimuló áreas ligadas a la memoria en el lóbulo temporal, porque en muchos casos llegó a reactivar distintos recuerdos de los pacientes, quienes parecían revivir su pasado como en una película. Por ejemplo, una paciente dijo que le venía el recuerdo de estar en la cocina de su casa escuchando la voz de su hijo que jugaba afuera y alegaba ser consciente de los ruidos de la vecindad, de los autos que pasaban y de cómo estos eran un peligro para su hijo. Al estimular un punto específico de la corteza temporal de otra paciente, esta reportó escuchar la voz de una mujer y un hombre que la llamaban en algún lugar a lo largo de un río, y estimulando en otro punto, refirió tener una memoria familiar de estar en una oficina con escritorios y un hombre apoyado en uno de ellos con un lápiz en la mano. Penfield argumentó que estos recuerdos estaban guardados en el lóbulo temporal y que la estimulación no hacía más que evocarlos disparando el flujo de la conciencia, así como el sabor de la famosa magdalena de Marcel Proust disparara los recuerdos de su infancia en *En busca del tiempo perdido*[63].

Quisiera también mencionar un resultado muy curioso de mi colega Itzhak Fried, el neurocirujano con quien trabajo hace años en el estudio de registros neuronales en pacientes epilépticos, quien, dicho sea de paso, también es un profundo admirador de Penfield. Itzhak mostró que la estimulación de la corteza motora suplementaria de un paciente disparaba su risa. Cuando le preguntaron (durante la operación) de qué se reía, el paciente argumentó que el equipo de cirujanos “se veía muy gracioso; todos alrededor suyo”[64].

En conclusión, los resultados obtenidos por diversos grupos de neurocientíficos muestran que a partir de la estimulación eléctrica pueden generarse distintas sensaciones y hasta incluso evocar memorias. La neurociencia parecería entonces respaldar la idea subyacente en *Matrix*. Sin embargo, veremos en breve que esto no es tan así.

La invención de Morel es sin duda la obra más lograda de Adolfo Bioy Casares. Jorge Luis Borges (amigo entrañable de Bioy) afirma en su

prólogo que no le parece una hipérbole calificar su trama de perfecta. Un fugitivo llega a una isla abandonada en donde solo hay un museo, una capilla y una pileta de natación. Un buen día, lo despierta la música de un fonógrafo y ve un grupo de parejas con ropa algo pasada de moda bailando en la terraza que da al museo. El fugitivo se esconde temiendo ser descubierto por los intrusos, entre ellos una mujer que se aleja del resto para ver la puesta del sol en las rocas. Tras días de observarla a escondidas, termina enamorándose perdidamente de ella. Finalmente decide hablarle, pero, muy a su pesar, esta ni siquiera se molesta en contestarle; continua mirando el atardecer, leyendo su libro, como si él no existiera. Asimismo, en distintas situaciones el fugitivo cree haber sido descubierto por los intrusos y espera que vayan a buscarlo a la playa. Sin embargo, estos parecen desestimar completamente su presencia, incluso tras haber quedado imprevistamente cara a cara con uno de ellos.

La resolución de Bioy Casares es sorprendente: los intrusos no son más que proyecciones, con un increíble realismo, de lo vivido por un grupo de amigos en la isla hace varios años. ¿Pero por qué me detuve a contar esta historia? Porque en los pormenores de su trama, tan alabada por Borges, se halla la solución a la Matrix y la concepción del “cerebro en un frasco”.

Para que el argumento de *La invención de Morel* funcione, es vital que el fugitivo no interactúe con los intrusos y que esta falta de interacción se dé como algo natural. La justificación de Bioy es perfecta: por un lado, el fugitivo, escapando de la ley, busca no ser descubierto y, por el otro, los intrusos, altaneros, ni siquiera se dignan a reconocer su presencia. Si la mujer que veía el atardecer le hubiera contestado, o si aquella otra persona que quedó súbitamente frente a él hubiera gritado del susto, el argumento se caería a pedazos. El motivo es muy simple: un grupo de personas irreal no puede reaccionar o interactuar con alguien de carne y hueso. Imaginemos por un momento que la mujer de los atardeceres contestara las palabras del fugitivo. ¿Qué debería decir? ¿Cómo podría la mente brillante que generó estas increíbles proyecciones, ya sea Morel, el genio maligno de Descartes o el arquitecto de la Matrix, anticipar la

pregunta del fugitivo para dar una respuesta acorde en el momento indicado? ¿Cómo podría prever si la reacción debería ser ante un insulto, una caricia, o una declaración de amor?

En las primeras páginas de su obra maestra, *Consciousness explained*, el filósofo Daniel Dennett hace una descripción brillante de este punto. Una realidad ilusoria podría funcionar siempre y cuando no haya interacción de aquel que es engañado con dicha realidad. A partir del momento en que se establece un intercambio de información, el engaño se vuelve insostenible.

Un hábil ilusionista podría, por ejemplo, convencernos de estar hablando en el escenario aunque su imagen no fuera más que la proyección de un sofisticado holograma. Para resolver el truco podría hacer que la imagen se desvaneciera de repente y —abracadabra— aparecer mezclado entre el público en uno de los palcos. Por supuesto, para que el truco funcione el ilusionista necesita que nadie lo interrumpa y, más aún, que a nadie se le ocurra subir al escenario. Cuanta más libertad de acción le dé a su audiencia, cuanto más interactúe con ellos antes de desaparecer —por ejemplo, aparentando tener una conversación casual con un miembro del público confabulado— más creíble será su imagen proyectada, pero más sofisticado (y espectacular) deberá ser el engaño, coordinando a la perfección dichas interacciones.

Volvamos a *Matrix*. ¿Cómo sé que el mundo exterior es real? Porque responde coherentemente a mis expectativas. El genio maligno de Descartes podría en principio tener mi cerebro en un frasco, haciéndome creer que estoy en una playa paradisíaca. Para ello debería estimular neuronas específicas en mi corteza visual y así inducirme la sensación de ver el mar azul, la arena, las palmeras y demás. Pero al tocar la arena debería también simular en mi corteza somatosensorial la sensación que genera esta escurriéndose entre mis dedos; al ir al mar, tendría que simular el cambio de temperatura y la sensación del agua en mi cuerpo, y si chapoteo en las olas, cada movimiento de mi mano debería producir sensaciones táctiles y una intrincadísima ilusión visual de espuma y miles de gotas al viento. En otras palabras, quizás sea posible mantener la ilusión mientras me

mantenga postrado en la arena, como en una pesadilla en la cual no puedo moverme (así como para el ilusionista es posible generar la ilusión de que una imagen ficticia es real mientras mantenga distante al público), pero a partir del momento en que interactúo con la ilusión, esta se hace técnicamente imposible de sostener, incluso para un genio poderoso y maligno. Dennett dice que algunos filósofos argumentan que, por más difícil que sea lograr tal ilusión, es en principio posible. A lo cual contesta, a mi entender acertadamente, que también "es en principio posible construir una escalera de acero inoxidable hasta la luna o escribir, en orden alfabético, todas las conversaciones posibles en inglés con menos de mil palabras" (*Consciousness explained*, p. 4).

El mismo argumento se aplica a la existencia de los otros. Es imposible simular una interacción coherente con alguien que no existe; es decir, especificando cómo y cuándo esta persona ilusoria debería reaccionar a cada una de mis acciones. La variedad de acciones posibles diverge exponencialmente a medida que trascurre la interacción, ya que, incluso arrancando del mismo tema, la más frugal de las conversaciones puede tomar una infinidad de caminos.

El golpe de gracia al solipsismo y la posibilidad de la Matrix fue dado por Ludwig Wittgenstein (1889-1951), considerado por muchos el exponente más brillante de la filosofía del siglo XX. La historia de Wittgenstein es tan peculiar que vale la pena dedicarle unas líneas. Nacido en una familia muy adinerada de Austria, Wittgenstein estudió ingeniería aeronáutica en Berlín y luego en Manchester. Sus estudios de ingeniería lo llevaron a interesarse en la matemática y sus fundamentos filosóficos, y eventualmente terminó siendo discípulo de Bertrand Russell en Cambridge. Tal fue la pasión que desarrolló por la filosofía, que solía recluirse por largos períodos a pensar sobre problemas filosóficos en una cabaña aislada que construyó él mismo en los fiordos de Noruega.

Al estallar la Primera Guerra Mundial, en 1914, Wittgenstein se alistó como voluntario en la artillería austriaca, en la que sirvió y fue condecorado por su coraje. Fue tomado prisionero de guerra y durante esa etapa terminó el único libro que publicaría en vida, el *Tractatus*

logico-philosophicus. En el *Tractatus*, el autor se plantea cómo es la naturaleza de las representaciones que hacemos de las cosas y concluye que tales representaciones están dadas por los pensamientos, que a su vez son como imágenes. Según Wittgenstein, cuando nos comunicamos intercambiamos imágenes. Por ejemplo, al decirle a una persona que me voy de vacaciones al Caribe, creo en ella una imagen de mí mismo tirado en una playa. A partir de esa idea, Wittgenstein concluye que muchos de los problemas en la comunicación se deben a malentendidos generados por la interpretación errónea de las imágenes que generamos en los otros. Asimismo, argumenta que la filosofía no hace más que enmarañar las cosas originando un sinfín de malentendidos y magistralmente dice al respecto, en la última oración de su libro: "Cuando no podemos hablar de algo, hay que guardar silencio".

Al volver de la guerra, Wittgenstein hereda la fortuna de su fallecido padre y se vuelve una de las personas más ricas de Europa. Sin embargo, renuncia a todo ese dinero y, creyendo haber resuelto los problemas más fundamentales de la filosofía, trabaja como maestro rural, jardinero y hasta hace las veces de arquitecto en la construcción de la casa de su hermana. Aunque tras unos diez años, comienza a dudar de sus principios y retoma la filosofía en Cambridge. El resultado de esta segunda etapa es un libro que se publica post mórtem, *Investigaciones filosóficas*, donde critica denodadamente sus ideas anteriores.

En el *Tractatus*, Wittgenstein veía a las palabras como representaciones; entidades objetivas y absolutas. En su obra póstuma cambia radicalmente de visión y argumenta que el significado de las palabras está dado por el contexto y su uso en el lenguaje. Tal variabilidad otorga a las proposiciones un valor relativo, dependiendo de la interpretación subjetiva que se da a las palabras según el contexto. Por lo tanto, no hay un criterio absoluto para afirmar la veracidad o falsedad de las proposiciones, y tales juicios se dan a partir de las interacciones con otras personas. Dicho de otra manera, a través del lenguaje nos ponemos de acuerdo en afirmar que algunas proposiciones son ciertas y que otras tantas son falsas. Así,

Wittgenstein refuta la posibilidad de un *lenguaje privado*; por ejemplo, podemos afirmar que el sol del atardecer es rojo porque todos nos pusimos de acuerdo en llamarlo de esa manera[65]. Dice Wittgenstein (*Investigaciones filosóficas*, 258):

> Imaginémonos este caso. Quiero llevar un diario sobre la repetición de una determinada sensación. Con ese fin la asocio con el signo "S" y en un calendario escribo este signo por cada día que tengo la sensación. […]
> Una definición sirve por cierto para establecer el significado de un signo. — Bien, esto ocurre precisamente al concentrar la atención; pues, por ese medio, me imprimo la conexión del signo con la sensación.— "Me la imprimo", no obstante, sólo puede querer decir: este proceso hace que yo me acuerde en el futuro de la conexión correcta. Pero en nuestro caso yo no tengo criterio alguno de corrección. Uno diría aquí: es correcto lo que en cualquier caso me parezca correcto. Y esto sólo quiere decir que aquí no puede hablarse de "correcto".

Básicamente, Wittgenstein afirma que usando un lenguaje privado podríamos establecer un vínculo entre una sensación y un símbolo. Sin embargo, al volver a vivir esta sensación, tendríamos que confiar en nuestra memoria para recordar el vínculo correcto. Pero entonces no hay manera de saber si cometimos un error. En otras palabras, al no poder contrastar nuestras creencias con las de otros, se hace imposible tener una noción de lo que es correcto —lo que parece serlo, termina siéndolo. Por lo tanto, las afirmaciones producidas por una persona solitaria no tendrían sentido. Mas esto no es solo un problema de comunicación o de errores en alguna que otra palabra, pues nuestro pensamiento se basa en el uso del lenguaje —desarrollado a lo largo de los siglos, a través de infinidad de interacciones sociales— para poder consolidar y estructurar conceptos. El hecho de que tengamos lenguaje y de que podamos pensar y evaluar proposiciones implica, entonces, la existencia de los otros; la inviabilidad del solipsismo.

Los argumentos de Dennett y de Wittgenstein demuelen la posibilidad de Matrix. Sin embargo, la concepción de la realidad como una construcción en nuestro cerebro va muy en línea con la visión que hoy día tenemos en neurociencia. No hay un genio maligno que nos engaña, la realidad exterior evidentemente existe, pero en nuestro

cerebro apenas tenemos una representación muy pobre y subjetiva de esta; una noción muy en línea con los pensamientos de Kant.

Repasemos las ideas que discutimos a partir de *Matrix*. Con su famosa alegoría de la caverna, Platón dudó de la información provista por los sentidos. Siguiendo esta línea de pensamiento, Descartes luego se preguntaría si la realidad exterior podría ser fruto de un monumental engaño por un genio maligno, lo cual eventualmente llevó a idealistas como Berkeley, conscientes de cuan fácilmente pueden engañarnos los sentidos, a negar su existencia. Estos aparentes engaños, sin embargo, no son más que los procesos que tiene que realizar el cerebro para interpretar la información provista por los sentidos.

A fines del siglo XIX, el notable físico y médico alemán Hermann von Helmholtz (1821-1894)[66] observó que la información captada con los sentidos es muy pobre y que, por lo tanto, el cerebro realiza *inferencias inconscientes* para dar sentido a la realidad exterior. En otras palabras, el cerebro interpreta datos incompletos haciendo predicciones sobre la base de experiencias pasadas[67]. Por ejemplo, a cada momento creemos ver en detalle todo lo que está frente a nosotros, pero en realidad vemos claramente no más que lo que ocupa la uña del pulgar con el brazo estirado; el resto lo asumimos[68].

Helmholtz argumentó que no vemos copias de la realidad, de los objetos en el exterior, sino *signos,* construcciones realizadas por nuestro cerebro. Estos signos no necesariamente tienen que asemejarse a la realidad; basta con que sean reproducibles. Es decir, no hace falta que la representación que hago de un objeto sea parecida al objeto mismo; alcanza con que sea la misma al ver el objeto en distintas oportunidades. Dice Helmholtz (*The facts of perception*, 1878):

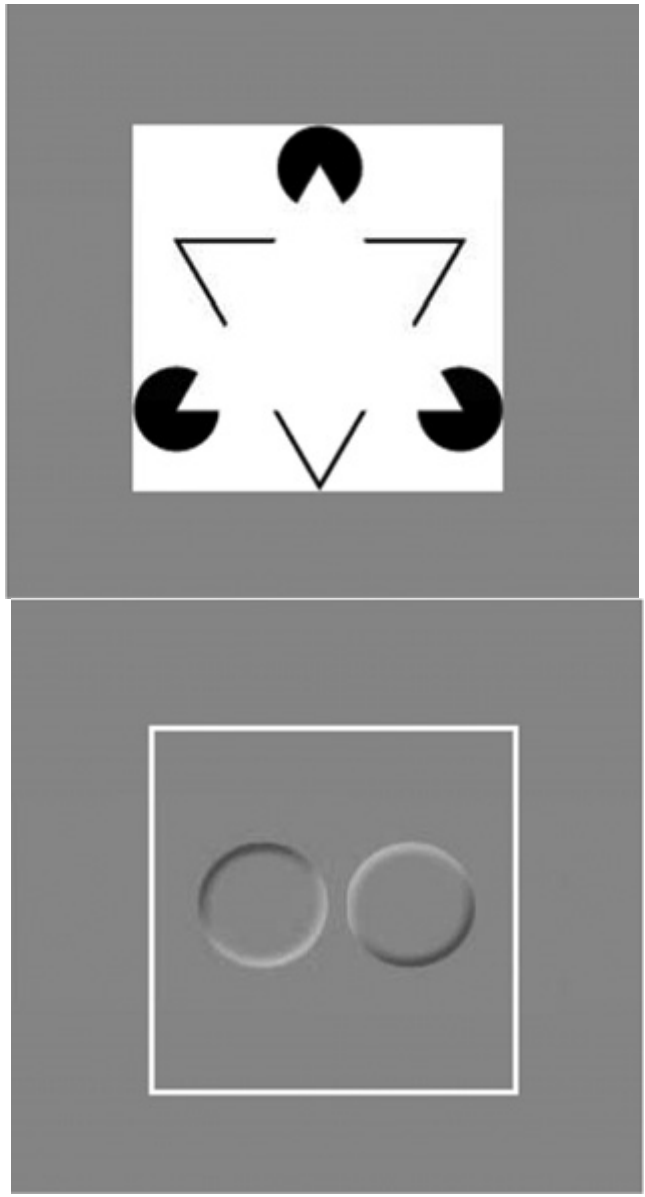
> [The objects in the space around us appear to possess the qualities of our sensations. They appear to be red or green, cold or warm, to have an odour or a taste, and so on. Yet these qualities of sensations belong only to our nervous system and do not extend at all into the space around us. Even when we know this, however, the illusion does not cease.]

> Los objetos que nos rodean poseen las cualidades de nuestras sensaciones.

Parecen ser rojos o verdes, fríos o calientes, parecen tener un olor y un sabor, y así sucesivamente. Sin embargo, estas cualidades de las sensaciones pertenecen sólo a nuestro sistema nervioso y no se extienden para nada al espacio a nuestro alrededor. Aun cuando sabemos esto, la ilusión no cesa.

La diferencia entre nuestra percepción de la realidad y la realidad misma es a todas luces evidente. Después de todo, los colores y los sonidos no existen en el espacio exterior, lo que existe son ondas electromagnéticas de distinta frecuencia impactando fotorreceptores en la retina, y variaciones en la presión en el aire que hacen vibrar pequeños filamentos en la cóclea. Los distintos colores, sonidos, olores, texturas y gustos son construcciones del cerebro; nuestra interpretación de la realidad exterior.

Las ilusiones visuales dan un claro ejemplo de cómo el cerebro interpreta la realidad a partir de inferencias inconscientes. La figura de abajo muestra una ilusión clásica, el triángulo de Kanizsa, en la que inferimos la forma de un triángulo a partir de sus ángulos y la aparente oclusión de otro triángulo en el fondo. Aun sabiendo que físicamente el triángulo no existe, la percepción de los lados del triángulo es inevitable. A la derecha vemos dos círculos en relieve, uno entrando y el otro saliendo del plano de la figura; pero son exactamente el mismo círculo rotado 180 grados. La ilusión de distintos relieves viene dada por las distintas sombras (en la parte superior e inferior, respectivamente), al asumir que la luz viene de arriba, como dicta nuestra experiencia.



Triángulo de Kanizsa y una ilusión de relieve.

Las inferencias descriptas por Helmholtz son las que nos llevan a percibir la realidad, a dar un sentido a lo que vemos. Tales inferencias son la base del funcionamiento del cerebro. Las usamos hasta en las cosas más mundanas, cuando jugamos al fútbol, cuando escuchamos música o cuando miramos una película. En un partido de fútbol, por ejemplo, vemos cómo se perfila un adversario, hacia dónde mira (y una infinidad de otros detalles de los que ni siquiera somos

conscientes en el momento) y predecimos si hará un pase corto, si tirará un centro, o un remate al arco. Cuando escuchamos una canción conocida, predecimos lo que viene basándonos en nuestra memoria, pero incluso cuando la escuchamos por primera vez también hacemos una infinidad de predicciones, ya que inconscientemente esperamos que la melodía siga ciertas reglas (usando notas de escalas específicas). Cuando vemos una película, se establece un juego entre el espectador y el realizador: el primero trata de adivinar la trama y el segundo busca que no sea muy evidente, aunque tampoco quiere violar demasiado las expectativas del espectador, pues sabe que una película completamente impredecible sería insoportable. A este respecto, Alfred Hitchcock, el maestro del cine de suspenso, decía que una explosión no causa terror, pero sí su anticipación. Según Hitchcock, genera mucha más tensión mostrar una bomba antes de que explote que hacerla explotar sorpresivamente.

Las inferencias de las que hablamos nos permiten predecir el futuro y anticiparnos a los hechos, lo cual no solo nos lleva a jugar bien al fútbol, disfrutar un tema de Pink Floyd o una película de Bergman; tales inferencias son de hecho vitales para nuestra supervivencia. Si veo una mudanza en la cual tratan de entrar un piano por el balcón con cuerdas afirmadas en la terraza, imagino lo que podría llegar a pasar y cruzo la calle; si veo un perro rabioso gruñirme amenazante, doy media vuelta.

Ya al final de la película, Neo llega a desentrañar como funciona Matrix; puede ver claramente el código que genera la ilusión del pasillo del hotel y los tres agentes que enfrenta en ese momento. A mi entender, esta escena da una analogía perfecta de cómo los magos ven nuestro comportamiento. Son los Neo del mundo real, dada la noción que tienen de cómo vivimos haciendo predicciones a partir de experiencias pasadas.

Hace unos años invite a Miguel Ángel Gea, un gran mago y amigo, a dar una charla sobre magia en un aula repleta en mi universidad. Comenzó su exposición diciendo que los presentes eran todos académicos, personas muy cultas e inteligentes y, consecuentemente… ¡muy fáciles de engañar! La reacción fue un coro de carcajadas y

aplausos, pero Gea hablaba muy en serio. Las personas inteligentes —continuó diciendo— constantemente hacen suposiciones sobre la realidad que los magos violan a gusto y placer en cada uno de sus trucos. De hecho, la magia para chicos es totalmente distinta a la magia para adultos, porque los chicos se detienen a mirar detalles que los adultos, con el tiempo, aprendimos a ignorar completamente.

Generalmente los magos tienden a no repetir el mismo truco. Esto es porque al no saber qué esperar, el espectador, bombardeado de información que no llega a procesar del todo, no sabe de antemano cuándo o hacia dónde enfocar su atención. Harry Houdini, quien fuera un escapista notable y quizás la personalidad más renombrada en el mundo de la magia a principios del siglo XX, propuso el siguiente desafío: si alguien le hacía un truco tres veces, cualquier truco, él iba a darse cuenta de cómo había sido hecho. En febrero de 1922, un joven mago llamado Dai Vernon —en aquel entonces totalmente desconocido, aunque después sería considerado una de las grandes figuras de la magia— pidió a Houdini que eligiera una carta cualquiera y la firmara. Luego introdujo la carta en el mazo, procedió a mezclar y, abracadabra, la carta elegida quedó arriba de todo. Perdió nuevamente la carta en el mazo, volvió a mezclar y la carta otra vez apareció arriba. Vernon repitió el truco siete veces y Houdini miraba asombrado sin tener la menor idea de cómo podía hacerlo. Lo que no sabía era que Vernon contantemente cambiaba el método con el cual hacía que la carta volviera a aparecer arriba de todo. Cada vez que Houdini sospechaba cómo había hecho el truco y esperaba confirmar esta sospecha al verlo de nuevo, Vernon volvía a cambiar el método. En otras palabras, lo que confundía a Houdini era el hecho de hacer la más que razonable suposición de que si el resultado era siempre el mismo —mezclaba y la carta volvía a aparecer arriba de todo— Vernon hacía el truco de la misma manera.

Harry Houdini (izquierda) y Dai Vernon (derecha), dos de los magos más célebres del siglo XX.

Para la mayoría de las personas, la magia se basa en principios como “la mano es más rápida que la vista” o en el hecho de que el mago hace que el espectador mire a otro lado (por ejemplo, provocando una distracción) cuando realiza el truco. Pero la manera en que los magos hacen sus trucos suele ser más interesante[69]. De hecho, muchos trucos con cartas no requieren gran velocidad o increíbles dotes de prestidigitación y se pueden hacer relativamente lento[70]. Por otro lado, distraer al público haciéndolo mirar hacia otro lado puede arruinar el efecto; esa sensación de no entender cómo una carta puede aparecer arriba de todo, pese a haber mantenido la vista en el mazo todo el tiempo. El ingrediente fundamental de la magia es justamente lo que Vernon usó para engañar a Houdini, aprovechándose de la muy pobre capacidad que tenemos de registrar y procesar información y de cómo compensamos esta limitación haciendo inferencias inevitables e inconscientes. Neurociencia pura y, en este caso, con un toque de filosofía. La inferencia que usó Vernon tiene sus raíces en los pensamientos de David Hume en el siglo XVIII, quien alegara que, si un efecto se repite una y otra vez, es muy difícil evitar inferir que se debe a la misma causa.

El mismo principio que Helmholtz describiera sobre la percepción también se aplica a la memoria. Frederic Bartlett (1886-1969), un filósofo y psicólogo inglés, hizo a leer a sus estudiantes de la

Universidad de Cambridge una historia tradicional de aborígenes norteamericanos (“La guerra de los espíritus”) y luego les pidió que se la repitieran tras días, semanas y, en algunos casos, hasta varios años después de haberla leído. Bartlett observó que las historias recapituladas tendían a ser mucho más cortas y simplificadas y que cada sujeto las modificaba de acuerdo con su interpretación personal[71]. Más aún, los sujetos tendían a cambiar la historia cada vez que la repetían y en algunos casos, tras varias repeticiones, la recapitulación terminaba teniendo poca semejanza con la historia original. En otras palabras, más que la historia en sí, los estudiantes recordaban el *esquema* que se habían hecho de esta según la interpretación que hicieran en el momento de leerla. A partir de este esquema reconstruían la historia de distintas maneras, olvidando muchos detalles e inventando y agregando inconscientemente otros tantos. Bartlett concluyó entonces que la memoria es un proceso creativo y que su consolidación afianza una representación subjetiva, la cual muchas veces cambia el recuerdo mismo.

Hay una similitud muy grande entre la construcción de *signos* de Helmholtz y la generación de *esquemas* de Bartlett. En un caso nos referimos a la visión y en el otro a la memoria, pero el proceso es esencialmente el mismo; implica la construcción de un significado de la realidad a partir de inferencias inconscientes, y el uso de ese significado, llámese signo o esquema, en vez de la realidad misma. Y así como las inferencias inconscientes dan origen a ilusiones visuales, análogas inferencias dan origen a falsas memorias; al afianzamiento de recuerdos que no se corresponden con la experiencia original.

Hugo Gatti, “el león de Kiev”, atrapando la pelota (blanca) en el partido de la Argentina contra la Unión Soviética.

Antes mencioné al pasar que para muchos argentinos los recuerdos más fuertes están relacionados con partidos de fútbol. Justamente una de las memorias más salientes de mi infancia es la de un partido entre la Argentina y la Unión Soviética jugado en el invierno de Kiev. Recuerdo particularmente ese partido por lo raro que era ver jugar al fútbol en la nieve. También tengo un recuerdo muy vívido de la pelota naranja. Pero, curiosamente, ese recuerdo es falso. Terminé afianzando en mi memoria algo que nunca pasó: la pelota era blanca.

La manera como descubrí que este recuerdo era una fabulación fue muy curiosa. Para sorpresa incluso de los más fanáticos, Leicester City —el equipo de fútbol de la ciudad donde vivo, que hasta entonces apenas lograba mantenerse en la primera división— salió campeón de la liga inglesa. Ezequiel Fernández Moores, brillante periodista deportivo, me llamó a Leicester para preguntarme cómo se vivía lo

que muchos en Inglaterra consideran el resultado más sorprendente de la historia del fútbol. Hablando de una y otra cosa, salió a la luz el recuerdo de la pelota naranja en el partido contra Rusia. Pero uno de los colegas de Ezequiel, con una memoria enciclopédica, me corrigió diciendo que la pelota era en realidad blanca y que, de hecho, existía una foto famosa de ese partido (en página anterior), en donde Hugo Gatti, "el león de Kiev", aparece tomándola en el aire. La confusión se generó porque tiempo después le pidieron al Loco Gatti que eligiera la pelota a usarse en un clásico entre Boca y River. Gatti asumió que la cancha iba a terminar repleta de papelitos y pidió usar una pelota naranja. El recuerdo de la pelota naranja quedó entonces consolidado en mi memoria, a partir de hacer una inferencia inconsciente que era incorrecta[72].

¿Por qué registramos y recordamos tan poco? Con billones de neuronas en el cerebro, ¿no podríamos procesar y guardar una representación mucho más fehaciente de la realidad? Esto se relaciona con la discusión de los capítulos anteriores, en donde comparábamos el funcionamiento de nuestro cerebro con el de las computadoras. La clave es que el cerebro no busca hacer una reproducción detallada de la realidad, sino que enfoca sus recursos en tratar de comprenderla. A diferencia de una computadora, que puede guardar miles de fotos, canciones y videos con altísima resolución, es realmente muy poco lo que vemos o recordamos. Aunque nosotros entendemos lo que nos pasa, la computadora no.

En "Funes el memorioso", Jorge Luis Borges describe las desventuras de un peón de Fray Bentos que a partir de recibir un golpe en la cabeza adquiere la increíble capacidad de recordar absolutamente todo. Dice que "Funes no solo recordaba cada hoja de cada árbol de cada monte, sino cada una de las veces que la había percibido o imaginado"[73]. Funes termina sus días postrado en la oscuridad de su habitación, abarrotado de memorias y detalles irrelevantes que no lo dejaban siquiera pensar. A partir de la historia de Funes, Borges muestra con increíble lucidez la importancia del olvido y el tormento que acarrearía recordarlo todo. De hecho, el autor concluye que Funes no era muy capaz de pensar, pues "pensar es

olvidar diferencias, es generalizar, abstraer”[74].

Nuevamente remarco que en un área del cerebro dedicada a la memoria descubrimos neuronas representando conceptos (Jennifer Aniston, Luke Skywalker, etcétera). Estas neuronas no codifican los detalles de una u otra persona —cómo están vestidos, si miran de frente o de perfil, etcétera— justamente porque tendemos a olvidar detalles y a recordar conceptos y asociaciones entre estos. A partir de estos conceptos generamos los esquemas de Bartlett, las representaciones subjetivas de cómo recordamos los eventos de nuestras vidas.

Vimos que la idea de *Matrix* es imposible, aunque nos ofrece una analogía muy útil de cómo funciona el cerebro, de cómo creamos una representación de la realidad. Si bien, por otro lado, el argumento de *Matrix* y la idea de estar inmersos en una gran simulación es de una actualidad escalofriante, considerando cómo vivimos conectados a Internet y cómo nuestra realidad del día a día está dada por un constante bombardeo de información, ya sea por el e-mail, WhatsApp, Messenger, cientos de canales de televisión, el celular o la tablet. Tal bombardeo no hace más que llevarnos a terminar como Funes, devorando tiempo valiosísimo que convendría dedicar a plantearnos infinidad de cosas. Asimismo, nuestras interacciones sociales dependen cada vez más de mecanismos de interacción virtuales como Facebook o Twitter y, como si fuéramos una de las tantas personas viviendo el engaño de *Matrix*, somos adictos a juegos de computadora y nos sentamos horas frente al televisor a vivir vidas ajenas: pasamos en minutos de ser héroes de acción a ser expertos en artes marciales, agentes secretos o ladrones de banco; a fin de cuentas, se nos va la vida frente a una u otra pantalla y somos cada vez menos nosotros mismos. Tomemos pues la píldora roja, apaguemos la computadora, desconectemos el celular y salgamos a la vida.

55 Neo tiene guardado el disco dentro de un libro titulado *Simulacra and Simulation*.

El tema recurrente de este libro, publicado en 1981 por el filósofo francés Jean Baudrillard, es la borrosa diferencia que existe entre lo real y lo ilusorio, aquello que es una simulación. Por ejemplo, Baudrillard contrasta lo que es presentado como ilusorio en Disneylandia y la (aparente) realidad de una ciudad como Los Ángeles (cualquiera que haya vivido un tiempo en Los Ángeles entiende perfectamente a qué se refiere). La confusa diferencia entre realidad e ilusión es el tema principal de *Matrix*. Los hermanos Wachowski, guionistas y directores de la película, de hecho, obligaron a los actores a leer el libro de Baudrillard antes de la filmación. La película también refiere claramente a *Alicia en el país de las maravillas*, al seguir Neo el conejo blanco.

56 Para Berkeley, así como *esse est percipi* (ser es ser percibido) se aplicaba a las cosas, *esse est percipere* (ser es percibir) se aplicaba a las mentes.

57 Para un tratamiento detallado del tema, ver, por ejemplo: Manuel García Morente, *Lecciones preliminares de filosofía* (Losada, 2004); Anthony Kenny, *Breve historia de la filosofía occidental* (Paidós, 2005); Bertrand Russell, *History of Western Philosophy* (Routledge Classics, [1946] 2004); John Hospers, *An introduction to philosophical analysis* (Routledge, 1956). La versión *online* de la *Stanford Encyclopedia of Philosophy* ofrece también una descripción muy detallada y clara del idealismo.

58 El obispo Berkeley también resuelve muchas de las contradicciones de su idealismo ontológico apelando a la existencia de Dios. Según Russell, si Berkeley hubiera sido consistente con su doctrina, tendría que haber negado conocimiento alguno sobre Dios y habría terminado en el solipsismo (Bertrand Russell, *History of Western Philosophy*, p. 564).

59 Hilary Putnam. *Reason, Truth and History* (1982). En base a argumentos lógicos que han sido muy discutidos en la filosofía contemporánea, Putnam afirmó que es imposible que seamos un *brain in a vat*, refutando de esta manera el escepticismo cartesiano en donde la realidad es una ilusión generada por un genio maligno. Para más detalles sobre los argumentos de Putnam ver el capítulo 1 de su libro. En las páginas siguientes presentaré, en cambio, una refutación mucho más sencilla y directa propuesta por Daniel Dennett y aquella hecha por Wittgenstein basada en el lenguaje.

60 Daniel Salzman, Kenneth Britten and William Newsome. “Cortical microstimulation influences perceptual judgement of motion direction.” *Nature* 346: 174-177; 1990.

61 Seyed-Reza Afraz, Roozbeh Kiani and Hossein Esteky. “Microstimulation of inferotemporal cortex influences face categorization.” *Nature* 442: 692-695; 2006.

62 Wilder Penfield; *The mystery of the mind*; Princeton University Press; 1975.

63 Sin embargo, la evocación de recuerdos en estos pacientes funcionó sólo en un 8% de los casos y en casi la mitad de estos, los recuerdos eran iguales a las alucinaciones que sufrían durante sus crisis epilépticas. Por lo tanto, puede también argumentarse que más que disparar el flujo de la conciencia, Penfield simplemente evocó crisis epilépticas. Pero más allá de estas posibles interpretaciones, el hecho de

que pudiera evocar recuerdos a partir de estimulaciones eléctricas en el cerebro es espectacular.

64 Para más detalles, ver: Itzhak Fried, Charles Wilson, Katherine McDonald and Erik Benke. “Electric current stimulates laughter”. *Nature* 391: 650; 1998.

65 Las manifestaciones de Wittgenstein no son nada triviales y han dado lugar a múltiples interpretaciones por distintos filósofos. Probablemente mi descripción sucinta no le haga justicia a la magnitud de sus ideas. Para un tratamiento más detallado ver, por ejemplo, el capítulo 22 de Anthony Kenny, *Breve historia de la filosofía occidental* (Paidós, 2005). O, aún mejor, como dice Dennett, uno puede ciertamente ignorar las interpretaciones de Wittgenstein por otros filósofos y leer en cambio sus textos originales (Dennett, *Kinds of minds*, p. 170).

66 La diversidad de contribuciones de Helmholtz es realmente impactante. Entre otras cosas, formuló el principio de la conservación de la energía y postuló la noción de energía libre en termodinámica; inventó el oftalmoscopio, midió la velocidad de conducción de los nervios, introdujo una descripción matemática de las vibraciones acústicas y sentó las bases de la actual teoría del color usando tres variables para caracterizarlo (tono, saturación y brillo).

67 Argumentos similares fueron también hechos anteriormente por Ptolomeo, Aristóteles, Alhazen y Descartes, entre otros. Para más detalles, ver el capítulo 9 de *Borges y la memoria*, op. cit.

68 Para una descripción más detallada y cuantitativa de cuánto vemos, ver el capítulo 2 de *Qué es la memoria*, op. cit. El capítulo 3 de este libro también describe los procesos que implementa el cerebro para procesar información visual en base a inferencias inconscientes.

69 Para más detalles, ver: Rodrigo Quian Quiroga. “Magic and cognitive neuroscience”. Op. cit.

70 Justamente uno de los magos más notables de la Argentina, René Lavand, usaba una y otra vez el latiguillo “no se puede hacer más lento”, al manipular las cartas con su única mano. Visitando el Magic Castle en Los Ángeles, la meca de la magia, me sorprendió muy gratamente ver un retrato enorme de Lavand en una de las salas principales. Allí me enteré de que aquel hombre que conocía de uno u otro programa de televisión en la Argentina era considerado una de las luminarias de la magia en el mundo.

71 Para una descripción más detallada de estos resultados, ver el capítulo 5 de Frederic Bartlett, *Remembering*, Cambridge University Press, 1932. El capítulo 7 también describe resultados análogos con otras historias.

72 De hecho, creo que no soy el único que tiene una falsa memoria de la pelota naranja en el partido de Argentina en la nieve. Buscando evidencia sobre esto, vi un reportaje al Beto Alonso (quien la rompiera en el partido contra Boca) en donde decía algo parecido, y otro en *Clarín* (del 7/2/2013) en el cual también se afirma que la pelota en el partido contra la Unión Soviética en Kiev era naranja. Sin embargo, la foto del “león de Kiev” atrapando la pelota blanca es irrefutable.

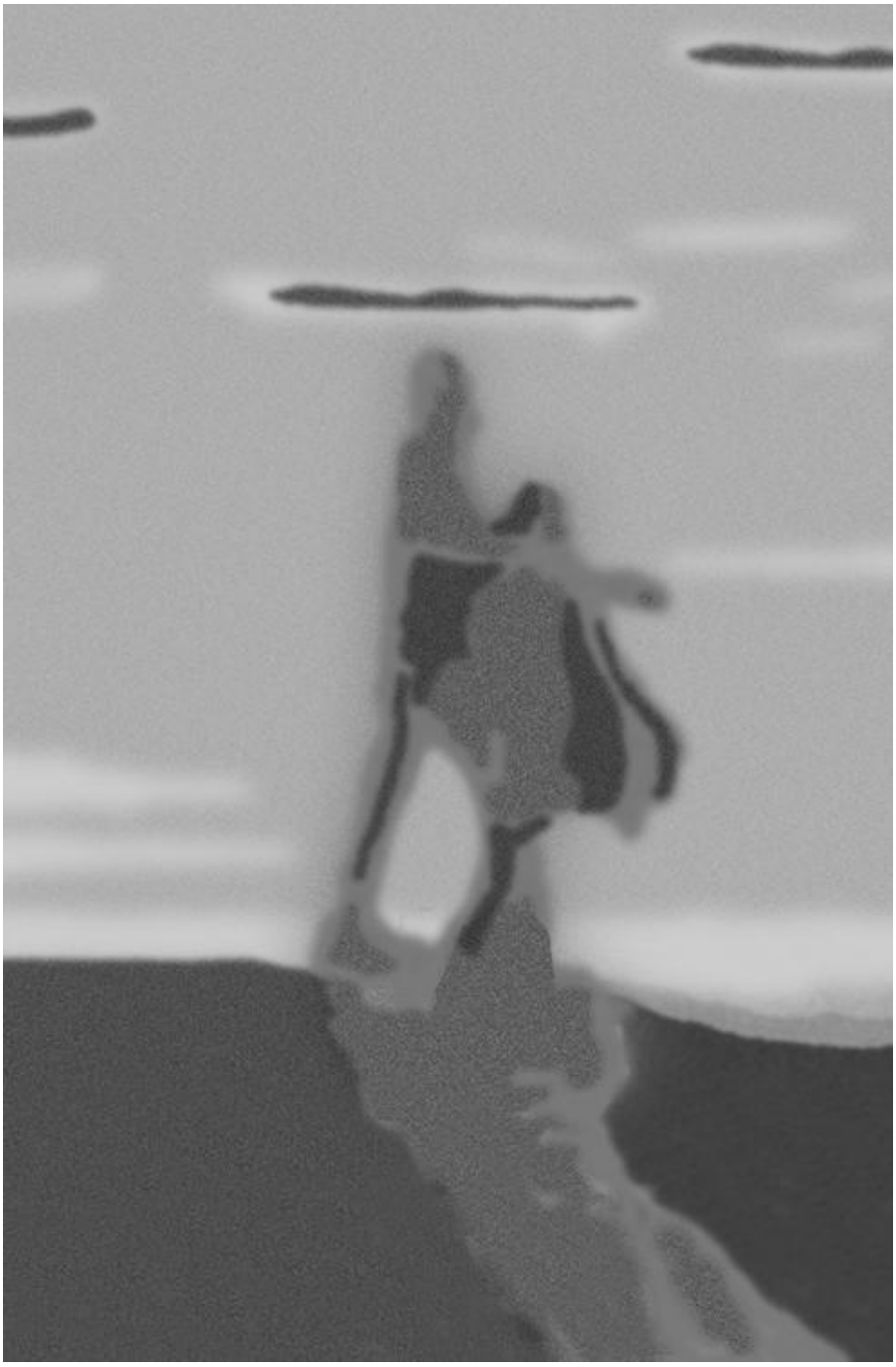
Para una descripción más detallada de falsas memorias, ver el capítulo 4 de *Qué es la memoria*, op. cit., o los trabajos de Elizabeth Loftus, una experta en el tema; entre ellos: Elizabeth Loftus. "Creating false memories". *Scientific American* 277: 70-75; 1997.

73 Jorge Luis Borges, *Ficciones*, op. cit.

74 La importancia del olvido en el pensamiento es el tema preponderante de *Borges y la memoria*, op. cit.

V. Hasta el fin del mundo

Leer la mente



Junto a Herzog y Fassbinder, Wim Wenders, director de *Hasta el fin del mundo*, es uno de los grandes exponentes de la nueva ola del cine alemán de los años 70. Reflejando su pasión por la fotografía y el rock, la película de Wenders combina música de U2, Lou Reed y Peter Gabriel, con paisajes imponentes en el desierto de Australia, o una alienación de formas y colores en el centro de Tokio. Es una especie de *road movie* filmada en diez países distintos, que remata con un final de ciencia ficción en un laboratorio aislado en el medio del desierto.

La mezcla argumental de la película fue unánimemente destrozada por la crítica, quizás en gran parte por las expectativas creadas por su película anterior, *Der Himmel über Berlin* (conocida en español como *Las alas del deseo*), filmada en una Berlín espectacular, un par de años antes de la caída del muro, en 1989.

Las alas del deseo es pura poesía, melancolía, filosofía y humanismo. Comienza con el poema "Lied Vom Kindsein" de Peter Handke, coescritor del guion junto a Wenders, que en la cuarta estrofa marca el tono y la profundidad de la película:

[Als das Kind Kind war,
war es die Zeit der folgenden Fragen:
Warum bin ich ich und warum nicht du?
Warum bin ich hier und warum nicht dort?
Wann begann die Zeit und wo endet der Raum?
Ist das Leben unter der Sonne nicht bloß ein Traum? […]
Wie kann es sein, daß ich, der ich bin,
bevor ich wurde, nicht war,
und daß einmal ich, der ich bin,
nicht mehr der ich bin, sein werde?]

Cuando el niño era niño,
era el tiempo de hacerse las siguientes preguntas:

¿Por qué soy yo, y por qué no tú?
¿Por qué estoy aquí y por qué no allá?
¿Cuándo comenzó el tiempo y dónde termina el espacio?
¿Es la vida bajo el sol no más que un sueño? […]
Cómo puede ser, que yo, lo que soy,
antes de ser yo, no fuera,
y que alguna vez yo, lo que soy,
no seré más lo que soy.

Dos ángeles, Damiel y Casiel, sobrevuelan Berlín acercándose a distintas personas, escuchando sus pensamientos y tratando de entender cómo viven y qué sienten. Sin la capacidad de sentir ellos mismos, se juntan cotidianamente a intercambiar experiencias de las que han sido testigos: la vida de niños, padres, abuelos, hombres a punto de morir, adultos desconsolados y otras tantas personas atrapadas en su rutina diaria. En gran parte la película es un continuo fluir del consciente —el consciente de aquellos cuya voz interna los ángeles sintonizan como con el dial de una radio— y contrasta la melancolía de la vida cotidiana de los humanos, con el eterno anhelo que tienen los ángeles de poder sentir por sí mismos.

En un juego estético brillante, Wenders combina escenas en color, representando el punto de vista humano, con otras en blanco y negro, mostrando la visión de la realidad de los ángeles, incapaces de ver o sentir colores. Damiel confiesa a Casiel que añora sentir lo que es llegar a su casa, sacarse los zapatos debajo de la mesa y estirar los dedos con los pies descalzos; comer cordero asado con vino, tener fiebre, los dedos negros de leer el diario; tener nada más que sospechas en vez de saberlo todo. Tal es su hastío, que llegado el punto decide resignar su inmortalidad para convertirse en un hombre más.

Columna de la Victoria (*Siegessäule*) en Berlín, donde dos ángeles, Damiel y Casiel, se encuentran para compartir sus experiencias con los humanos en *Las alas del deseo*.

Al despertar, ya humano, camina exultante y siente el peso de su cuerpo en cada paso; se cruza con un transeúnte y le pregunta por los distintos colores en la calle, pues aunque desde el principio de los tiempos sabe perfectamente qué son los colores y todo lo que representan, finalmente puede llegar a verlos. Pide un café, toma la taza entre las manos, siente su calor y, por primera vez, se detiene a oler el aroma tan particular del café antes de saborearlo.

La historia de Damiel tiene mucho en común con la de un arcángel imaginado por el filósofo británico Charlie Dunbar Broad[75]. Este arcángel —según Broad— posee un conocimiento de química ilimitado y sabe, por ejemplo, todos y cada uno de los detalles de la composición microscópica del amoniaco y como este se comporta en infinidad de reacciones químicas. Pero así como Damiel no conocía el olor del café antes de volverse humano, Broad argumenta que su arcángel, más allá de todo el conocimiento de química que pueda tener, no llegaría siquiera a imaginar cómo es el olor penetrante del amoniaco. En otras palabras, el conocimiento de procesos materiales no puede explicar cómo sentimos la realidad.

Sigamos un poco más con *Las alas del deseo* porque nos lleva a refrescar la discusión sobre el materialismo y, a partir de allí, la posibilidad de leer la mente. En la narrativa visual de Wenders dominan los planos nostálgicos de una Berlín en blanco y negro; los colores mayormente aparecen recién sobre el final de la película, cuando Damiel se vuelve humano. Esto justamente rememora otra gran discusión en la filosofía contemporánea: la historia de María, una científica imaginada por el filósofo australiano Frank Jackson[76].

María es una experta en percepción visual que pasa su vida encerrada en un laboratorio donde absolutamente todo es blanco o negro: el instrumental, los libros, el monitor de su computadora, la televisión. Como Damiel, María nunca vio colores, aunque entiende perfectamente lo que significan y las respuestas que estos generan en el cerebro; María sabe todo lo que hay que saber sobre la percepción de los colores. Pero un día sale de su habitación y, así como Damiel al volverse humano, ve colores por primera vez. ¿Se sorprenderá entonces al ver los colores? Si ya sabe todo lo que hay por saber sobre ellos, ¿qué información agregaría el poder sentirlos?

Esto es lo que se conoce como el *argumento del conocimiento* para refutar el materialismo. Al salir de la habitación, María estaría aprendiendo algo nuevo al ver por primera vez el rojo de una manzana. Pero si ya sabía absolutamente todo sobre los procesos físicos que se desencadenan al ver los distintos colores (es decir, las respuestas neuronales que estos generan) y si siente algo nuevo al ver

colores por primera vez, entonces los procesos mentales deben involucrar información que va más allá de lo material. Por lo tanto —concluye Jackson— la mente no es reducible a procesos físicos y el materialismo es falso.

Argumentos similares fueron hechos anteriormente por el filósofo Thomas Nagel al tratar de imaginar cómo debería sentirse *ser* un murciélago[77]. Esto dista mucho de imaginarse a alguien como Drácula. El planteo de Nagel es muy distinto: no se pregunta cómo se sentiría un hombre viviendo como un murciélago, sino qué se siente ser murciélago; cómo funciona la percepción que tienen de la realidad y cómo es su pensamiento.

La elección del murciélago no es caprichosa, ya que estos tienen una visión muy limitada y perciben sus alrededores, por ejemplo, en una caverna totalmente a oscuras, en función del eco generado por sonidos que emiten. Nagel dice que por más que entendamos todos los detalles de cómo funciona el sistema de ecolocalización de los murciélagos, nunca podríamos imaginar lo que se siente ser uno de ellos. Pero a diferencia de Jackson, Nagel no usa este argumento para refutar explícitamente al materialismo, sino para ilustrar la dificultad que tenemos de entender la relación entre los procesos físicos (es decir, el conocimiento de cómo funciona la ecolocalización) y los mentales (es decir, lo que sentimos al percibir los ecos). Esto es justamente el problema difícil del que habláramos anteriormente.

Volvamos a María y a la aparente refutación del materialismo propuesta por Jackson. Para Dan Dennett, el brillante filósofo que ya introdujéramos, el argumento de Jackson se basa en un error de intuición. Si uno sigue exactamente las premisas de Jackson —argumenta Dennett— entonces no hay problema alguno[78]. La falla radica en el hecho de que es difícil intuir lo que realmente significa decir que María sabe *absolutamente todo* sobre la percepción de colores. Si efectivamente supiera todo, entonces sabría cómo las neuronas en su corteza visual se activan y cómo, a su vez, transmiten esta información a cientos de otras áreas, disparando memorias, emociones y una infinidad de procesos conscientes e inconscientes. En otras palabras, María sabría de antemano cómo reaccionaría al

percibir el rojo de la manzana y al verla por primera vez no debería sorprenderse en absoluto, pues su reacción sería justamente la que esperaba, una acumulación de procesos que dan lugar a la *qualia*; las sensaciones que nos genera ver el rojo de una manzana, el marrón de la borra del café o el blanco de un jazmín.

Dennett argumenta que tenemos un problema similar al tratar de imaginar cómo siente un murciélago[79]. Nuevamente es una falla de intuición y no un problema filosófico. En este caso el problema viene dado por la dificultad que tenemos de imaginar cómo percibir el exterior a partir de un sistema de ecolocalización, interpretando todo lo que acontece a nuestro alrededor como con un sonar. Asumamos, por ejemplo, que tal sistema tiene una resolución de unos pocos metros en los murciélagos. ¿Qué sentiríamos entonces, siendo murciélagos, al ser atacados por un águila? Hasta el último par de segundos, absolutamente nada, porque ni siquiera distinguiríamos el peligro del águila que se acerca. ¿Y qué sentiríamos en el último par de segundos? Probablemente esto dependa de nuestra capacidad de establecer, con nuestro sistema de sonar, si el águila se dirige a nosotros, si tiene sus garras desplegadas en ataque, etcétera. A partir de estos argumentos, Dennett concluye que por más antiintuitivo que resulte, al adquirir información del funcionamiento del cerebro de un murciélago, podemos efectivamente comenzar a entender qué se siente ser uno de ellos.

En conclusión, el materialismo resiste firme los embates filosóficos de murciélagos y científicos incapaces de ver colores. Tras identificar y dejar de lado intuiciones engañosas, queda claro que no hay nada mágico o misterioso que dé lugar a nuestra mente; es simplemente la actividad de billones de neuronas. Ahora sí, nos planteamos entonces la gran pregunta de este capítulo: Si la mente no es más que el funcionamiento del cerebro, ¿podremos leer la mente? Es decir, ¿podremos predecir lo que está viendo, sintiendo o pensando una persona a partir de medir la actividad de sus neuronas? Obviamente no escucharíamos, como Damiel y Casiel, la mente de aquellos a nuestro alrededor con solo prestarles atención. Deberíamos, en cambio, usar mediciones muy sofisticadas de la actividad del cerebro y

complejos algoritmos para interpretar lo que dicha actividad representa. Este es justamente el argumento de *Hasta el fin del mundo*.

El arte es muy subjetivo. Como mencionara al comienzo del capítulo, *Hasta el fin del mundo* fue unánimemente destrozada por la crítica. Aunque en defensa de Wenders basta decir que en su versión final la película duraba cuatro horas y media, pero dado que era imposible de comercializar fue reducida a dos horas y media, como exigía el contrato con la productora. A pesar de la falta de una cierta coherencia narrativa, *Hasta el fin del mundo* sigue siendo una de mis películas favoritas. Quizás por la banda de sonido, quizás también por los planos espectaculares, pero creo, sobre todo, por la riqueza del argumento final.

El filme podría dividirse en dos partes, tal vez algo desconectadas. La primera es una *road movie* en donde un hombre va casi frenéticamente de uno a otro lado del mundo con una cámara especial que registra no sólo las imágenes que filma, sino también sus reacciones cerebrales al filmarlas. En la segunda parte, llega a un lugar remoto en el desierto de Australia. Allí se encuentra con su padre, un científico que ha montado un imponente laboratorio con el fin exclusivo de hacer ver a su mujer ciega. Para ello, planea transmitirle directamente al cerebro las señales cerebrales que se registraron durante la filmación de las distintas escenas. El proceso es muy agotador, pero tras infinidad de ajustes, de a poco la mujer comienza a reconocer escenas borrosas (como aquella que ilustra este capítulo) a partir de las señales que recibe. Mas una vez logrado esto, de por sí una proeza, el científico planea realizar el proceso inverso: en vez de implantar en el cerebro las imágenes registradas por la cámara, trata de interpretar los patrones de activación que el cerebro produce naturalmente. Y es así como empieza a leer la mente a partir de *decodificar* la información provista por los registros cerebrales, proyectando los pensamientos en una pantalla y llegando hasta a vislumbrar los pensamientos durante el sueño. La trama por supuesto carece de un sustrato científico; a fin de cuentas, no es más que ficción, aunque, sorprendentemente, ya no tanto.

Casi veinte años después de haber visto por primera vez la película,

quiso el destino que terminara haciendo algo parecido: sobre la base de las repuestas a Jennifer Aniston, Halle Berry y otras personas (que describiera en los capítulos anteriores), me pregunté si, como aquel científico de *Hasta el fin del mundo,* podría llegar a predecir lo que veía el paciente. (Recordemos que en pacientes epilépticos, una vez implantados electrodos intracraneales para localizar el foco epiléptico, registramos neuronas que respondían a fotos de distintas personas.) La idea es sencilla: si la neurona de Jennifer Aniston dispara, predeciré que el paciente está viendo una foto de Jennifer Aniston; si dispara la de Halle Berry, predeciré, en cambio, que está viendo a Halle Berry, etcétera. Al considerar más neuronas y cosas a predecir, la tarea se vuelve obviamente más compleja, pero podemos resolverla usando algoritmos de decodificación. Dejando de lado detalles técnicos, el algoritmo de decodificación predice cuál es la foto que da lugar a la respuesta observada en el grupo de neuronas[80]. El resultado fue que efectivamente pudimos predecir, mucho mejor que al azar, qué foto estaba viendo el paciente en cada momento[81].

Tiempo después nos planteamos una pregunta aún más fascinante: si podemos predecir lo que está viendo el paciente, ¿podremos también proyectar lo que piensa en una pantalla, como en *Hasta el fin del mundo*? Por supuesto necesitamos condiciones controladas. No podemos predecir y luego proyectar cualquier pensamiento, pero sí aquellos que involucren las personas para las cuales tenemos neuronas que responden a ellas. Por ejemplo, si vemos que dispara la neurona de Jennifer Aniston, proyectamos en el monitor su imagen; si vemos que dispara la de Marilyn Monroe, proyectamos en cambio la imagen de esta, y así sucesivamente. ¿Y cómo hacemos para que dispare una u otra neurona? Muy fácil, le pedimos al paciente que piense, de a una a la vez, en las personas a las cuales responden estas neuronas. Luego usamos algoritmos matemáticos de decodificación para predecir qué está pensando el paciente y finalmente proyectamos la imagen de aquello que pensaba en el monitor de computadora. Y así como el científico de *Hasta el fin del mundo* apenas contenía su emoción al ver cómo, poco a poco, aparecían imágenes del cerebro proyectadas en una pantalla, tiempo después, en un laboratorio real, terminamos

fascinándonos nosotros mismos al ver que los pacientes controlaban voluntariamente el disparo de sus neuronas y proyectaban sus pensamientos en la pantalla de una computadora[82].

Aclaro que no hicimos más que proyectar aquello que los pacientes quisieron mostrarnos voluntariamente. Esto requirió de cierto esfuerzo y concentración por parte de ellos, pensando en una u otra persona para que nuestros algoritmos matemáticos pudieran predecir y luego proyectar la foto de estas personas en la pantalla. Estamos muy lejos de hurgar en sus pensamientos privados. Por el contrario, podríamos afirmar que desarrollamos un método alternativo de comunicación: en lugar de decirnos lo que pensaban, los pacientes voluntariamente proyectaban los pensamientos. ¿Y por qué esto es importante? Primero, porque muestra la capacidad que tiene una persona de modificar a voluntad la actividad de sus neuronas a través del pensamiento; nuevamente rechazamos el dualismo cartesiano, mente y neuronas son una misma cosa. Segundo, porque esta tecnología podría eventualmente usarse en pacientes con *síndrome de enclaustramiento* (que sufren la condena de estar encerrados en sí mismos al no poder comunicarse ni realizar movimiento alguno), u otros con severos déficits motores (como el famoso físico Stephen Hawking), quienes tienen una limitada capacidad de movimiento y por ende de comunicación, ya sea para hablar, escribir o siquiera mover un dedo. Es posible entonces que en algunos años estos pacientes puedan tal vez comunicarse directamente con el mundo exterior a partir de la actividad de sus neuronas.

Un caso particularmente dramático es el de los pacientes en estado vegetativo. El estado vegetativo suele ser confundido con el de coma, pero no son lo mismo. Mientras que un paciente en coma está dormido e inconsciente, uno en estado vegetativo puede estar despierto, gesticular, abrir los ojos y aparentar mirar alrededor, aunque en realidad no mira ni siente nada, ya que no es consciente.

La característica principal del estado vegetativo es que el paciente no responde voluntariamente a comando alguno (por ejemplo, al pedirle apretar la mano, pestañear, mover voluntariamente los ojos, etcétera). En algunos casos estos pacientes se recuperan a los pocos

días o semanas, pero a veces se mantienen en ese estado durante años, alimentados artificialmente (en general pueden respirar por sus propios medios) y con ínfimas chances de recuperación. Es particularmente en estos casos, tras años de aislamiento, que uno se pregunta si realmente no son conscientes de nada, o si tal vez lo son pero, así como las personas con síndrome de enclaustramiento —con una muy limitada capacidad de movimiento y por ende de comunicación— no pueden realizar el más mínimo movimiento para demostrarlo.

Imaginemos por un instante la desesperación de una persona postrada en la cama que es consciente de lo que sucede a su alrededor, pero que no puede hacérselo saber a nadie. ¿Habrá algo específico que quiera decirnos? ¿Querrá quizás escuchar música, que apaguemos la luz o que no hagamos tanto ruido? ¿Sentirá dolor y querrá pedir a gritos que le aumenten la dosis de analgésicos? (Esto recuerda a la película *Awake*, en donde un paciente descubre que tras ser anestesiado aún sigue consciente y sufre el dolor de todo el procedimiento quirúrgico sin la posibilidad de mover siquiera un músculo para que los médicos lo noten). Tales consideraciones llevan también a plantearnos preguntas mucho más dramáticas: ¿querrá una persona seguir viviendo tras varios años en tales condiciones, considerando las prácticamente nulas posibilidades de recuperarse?

Seguramente el más mentado de estos casos es el de Terry Schiavo, quien sufrió un paro cardíaco en febrero de 1990. Los paramédicos pudieron revivirla, pero debido a una prolongada falta de irrigación sanguínea al cerebro, quedó inconsciente y unos meses después se dictaminó que estaba en permanente estado vegetativo.

En 1998, tras años en el hospital sobreviviendo con alimentación asistida y sin ningún tipo de respuesta que mostrara un mínimo grado de consciencia, su marido pidió que la desconectaran y la dejaran morir, dado que ese sería el deseo de ella si fuera consciente de su situación. Tal pedido fue rechazado por los padres de Schiavo, lo cual originó una extensa disputa legal, involucrando a distintos miembros del Congreso, la Corte Suprema de Justicia, la intervención del entonces gobernador de Florida, Jeb Bush, y hasta de su hermano,

George Bush, presidente de los Estados Unidos. Recién en marzo del 2005, tras siete años de múltiples dictámenes judiciales y apelaciones, Terry Schiavo falleció al desconectarse el tubo de alimentación que la mantenía viva.

El periplo legal desatado por el caso Schiavo abrió una fuerte disputa en la opinión pública. ¿Qué hacer en estos casos? ¿Es justificado usar métodos artificiales para prolongar durante años la vida de una persona que ya no tiene chances de volver a ser consciente? En gran parte, la disputa legal por Terry Schiavo se basaba en interpretar lo que ella habría deseado en esas circunstancias: mientras que su marido decía que hubiera querido irse en paz, sus padres afirmaban que, como ferviente católica que era, habría optado por continuar su vida, aun postrada como estaba. Suponiendo que todavía pudiera ser consciente, nos preguntamos entonces si de alguna manera seríamos capaces de leer su mente e interpretar realmente su voluntad. De hecho, la decisión judicial sobre su destino fue en parte demorada al creer los jueces que sería posible comunicarse con ella a partir de nuevos avances de la ciencia.

En septiembre de 2006, a poco más de un año del fallecimiento de Schiavo, un reporte en la revista *Science* generó todavía más revuelo[83]. El trabajo estuvo liderado por Adrian Owen, un neurocientífico británico que trabaja con casos como el de Terry Schiavo. Su labor es hercúlea y, en principio, frustrante: busca desarrollar métodos sofisticados para determinar si estos pacientes aún son conscientes, y de ser así, quebrar la barrera de su total postración e inmovilidad para eventualmente llegar a comunicarse con ellos. Para eso se vale de resonancia magnética por imágenes —un método que genera imágenes del cerebro y muestra sus activaciones al realizar distintas tareas— explotando el principio de que tanto el hecho de ejecutar como el de imaginar ciertas tareas genera activaciones estereotípicas en el cerebro: por ejemplo, la ejecución o la imaginación de un movimiento activa áreas involucradas en el procesamiento de tareas motoras.

El paciente analizado por Owen y sus colegas fue una mujer que había quedado en estado vegetativo tras un accidente automovilístico.

La mujer estuvo varios meses sin responder voluntariamente a ningún tipo de comando, sin manifestar signo alguno de estar consciente. Sin embargo, cuando Owen le pidió que imaginara estar jugando un partido de tenis mientras realizaba una resonancia magnética por imágenes, observó una clara activación en un área motora de la corteza cerebral, y cuando le pidió imaginar estar caminando por su casa, yendo de una a otra habitación, vio que se activaban otras áreas relacionadas con la navegación espacial, las mismas áreas que se activan normalmente en otras personas.

El siguiente paso es obvio, pero no por ello menos espectacular. Si el acto de imaginar distintas tareas genera activaciones específicas en el cerebro, esto podría usarse como un método de comunicación con los pacientes en estado vegetativo: simplemente se les pide que imaginen jugar al tenis cuando quieren decir que *sí* o que imaginen caminar por la casa cuando la respuesta es *no*. Esto es justamente lo que mostraron Owen y sus colaboradores en un trabajo publicado pocos años después[84]. Con un paciente (denominado paciente 23) que hacía cinco años permanecía en estado vegetativo, lograron mantener el siguiente diálogo:

Científico: ¿Tu padre se llama Thomas?
Paciente 23: No.
Científico: ¿Tu padre se llama Alexander?
Paciente 23: Sí.
Científico: ¿Tienes hermanos?
Paciente 23: Sí.
Científico: ¿Tienes hermanas?
Paciente 23: No.

Tras años sin poder comunicarse con nadie, el paciente 23 respondió correctamente cinco preguntas consecutivas. ¿Y por qué no siguieron preguntando? Porque para obtener respuestas confiables hace falta que el paciente se concentre en imaginar la acción de la respuesta durante alrededor de medio minuto y, además, necesitaban repetir la pregunta varias veces para validar la respuesta. Obviamente tanto Owen como los familiares del paciente hubieran querido tener

un diálogo más fluido y hacer infinidad de preguntas. Aunque si bien es cierto que la interacción es hasta ahora limitada, estos resultados son una clara demostración de que es posible leer la mente de pacientes en estado vegetativo.

La capacidad de realizar más preguntas o de tener interacciones más elaboradas es mayormente una cuestión de tecnología, optimizando los algoritmos de procesamiento de las señales, el tipo de preguntas que se hacen, etcétera. La mala noticia es que solo lograron comunicarse con alrededor de un 20% de estos pacientes. Quizás con los avances de la tecnología en el futuro sea posible aumentar ese número, pero ya el hecho de poder leer el cerebro de uno de cada cinco pacientes en estado vegetativo es de por sí una proeza. Aunque también es importante aclarar, para no generar falsas expectativas, que este procedimiento no es aplicable a todos los pacientes en coma, ya que en muchos casos las lesiones, involucrando estructuras primarias y vitales del cerebro, son irreversibles y no hay posibilidad alguna de comunicarse con la persona o de esperar que esta pueda eventualmente recuperar la consciencia.

Queda pendiente la gran pregunta: ¿querrán estos pacientes seguir viviendo en tales condiciones? Owen argumenta (a mi entender, correctamente) que todavía falta mucho como para plantear este tipo de preguntas. No solo es una cuestión de tecnología y de certeza en la interpretación de las respuestas, sino también de pensar qué hacer si eventualmente responden que no.

Más allá de su relevancia clínica, dejamos de lado el caso extremo de los pacientes en estado vegetativo, en quienes la posibilidad de leer la mente es crítica, aunque relativamente limitada, para preguntamos si podríamos extraer información mucho más sofisticada al implementar métodos análogos en sujetos normales.

Ya vimos que a partir de la implantación de electrodos en pacientes epilépticos podemos registrar la actividad de decenas de neuronas y predecir lo que están pensando e incluso proyectar sus pensamientos. La pregunta, entonces, es si será posible hacer algo análogo con métodos que no impliquen una cirugía y la implantación de electrodos en el cerebro. Por supuesto, la señal cerebral medida desde afuera de

la cabeza (sin intervención quirúrgica alguna), es muchísimo más débil e imprecisa que aquella obtenida al registrar directamente la actividad de las neuronas con electrodos intracraneales. Aunque tal vez el desarrollo de métodos más elaborados para el procesamiento de las señales permitan ir mucho más allá de obtener solo respuestas afirmativas o negativas como con los pacientes en estado vegetativo. En el fondo, no nos contentamos con obtener respuestas a preguntas predeterminadas y, así como el científico de *Hasta el fin del mundo*, nos preguntamos si será posible acceder directamente a los pensamientos de una persona.

Un grupo de investigadores israelíes liderados por Uri Hasson encontró algo sorprendente en este sentido[85]. Mientras mostraban a un grupo de sujetos la película *The good, the bad and the ugly*, escanearon su actividad cerebral con resonancia por imágenes (el mismo tipo de registros usados por Owen con pacientes en estado vegetativo) y encontraron que ciertas escenas de la película generaban las mismas activaciones específicas en todos los sujetos: los planos con caras activaban un área llamada *giro fusiforme*, aquellos con paisajes activaban un área aledaña y las escenas mostrando una manipulación de objetos con la mano activaban un área más distante.

Las activaciones de estas tres áreas del cerebro permiten entonces predecir si la persona está viendo caras, movimientos de manos o paisajes. La gran diferencia con los resultados obtenidos con pacientes en estado vegetativo es que en este caso no hubo que pedirles a los sujetos que se concentren durante medio minuto (y varias veces) en una acción determinada, sino que las activaciones, marcando uno u otro tipo de escena, se produjeron espontáneamente, a medida que los sujetos miraban la película.

Al considerar la posibilidad de leer la mente, la limitación más importante de estos trabajos es que la cantidad de áreas que se activa selectivamente viendo distinto tipo de imágenes, o haciendo determinadas tareas, es bastante limitada. Owen y su equipo fueron lo suficientemente inteligentes como para elegir tareas —jugar al tenis o caminar por la casa— que activan distintas áreas del cerebro. De la misma manera, el grupo de Hasson distinguió caras, de paisajes y

escenas con movimientos de manos. Pero ¿cuántas más cosas podrán identificarse a partir de la activación de áreas cerebrales específicas?

Así como las imágenes que vemos en el monitor de una computadora están formadas por píxeles, aquellas obtenidas con resonancia magnética por imágenes están compuestas por *voxels*. Para simplificar, podemos tomar a cada *voxel* como un píxel, cuya intensidad representa la activación de un pedacito de cerebro de un par de milímetros de lado. Grupos de *voxels* contiguos componen las áreas de las que hablamos anteriormente, con una extensión del orden de centímetros. Los trabajos del grupo de Owen y de Hasson, como otros tantos en neurociencia, justamente se basan en cuantificar la activación media del grupo de *voxels* que forman parte de distintas áreas.

El análisis de las activaciones promedio en distintas áreas ha sido el método clásico usado para la interpretación de las resonancias magnéticas por imágenes. Mas una gran revolución en el análisis de estos datos se dio a comienzos del siglo XXI, cuando el científico norteamericano Jim Haxby, junto a un grupo de colaboradores, se preguntó si la representación de caras y de diversos objetos está realmente dada por la activación de áreas específicas en la corteza visual[86]. El experimento consistió en mostrar fotos de caras, gatos y distintos objetos (casas, sillas, tijeras, zapatos y botellas) mientras registraban la activación de *voxels* en la corteza visual. Al analizar los resultados, Haxby y sus colegas no encontraron un área que respondiera a zapatos, u otra que se activara con diferentes tipos de botellas, sino que vieron que los distintos estímulos activaban las mismas áreas visuales, aunque no de la misma manera. Esto es lo que técnicamente se conoce como una representación *distribuida* y *solapada* de los estímulos. Sin embargo, a pesar de no contar con activaciones de áreas específicas, sobre la base del patrón de activación de los *voxels* en áreas visuales, pudieron predecir con un muy alto grado de certeza los estímulos que veían los sujetos.

La gran revolución vino dada por la metodología que desarrollaron para llegar a este resultado. Los detalles del método no son tan relevantes; lo importante es entender el hecho de que analizaron el

patrón de activación de cada uno de los *voxels* en las áreas visuales, en lugar de la activación media de una u otra área, como venía haciéndose hasta entonces. Vale aclarar que estos resultados no niegan que haya áreas que se activen preferentemente con caras o paisajes, como encontró el grupo de Hasson. Por el contrario, muestran que más allá de estas respuestas a gran escala, involucrando la activación conjunta de *voxels* en áreas determinadas, también existen activaciones a una escala mucho más fina, que pueden observarse al considerar cada uno de los *voxels* por separado y no solo como parte de algún área en particular[87].

Hace varios años me invitaron a dar una charla en una conferencia que se llevó a cabo en el estado de Maine, en los Estados Unidos. No tengo muy presente el tema de mi contribución, pero sí recuerdo vívidamente el impacto que me produjo la charla de Jack Gallant, un neurocientífico de la Universidad de California, en Berkeley. El grupo de Gallant ya venía obteniendo resultados sorprendentes, refinando el método desarrollado por Haxby para extraer información a partir del patrón de activación de los *voxels,* por ejemplo, prediciendo cuál de 120 imágenes veía un sujeto[88]. Pero los resultados que presentó en Maine fueron aún mucho más espectaculares. En este caso, primero registraron la actividad del cerebro —de los miles de *voxels*— mientras los sujetos miraban horas de videos: escenas de películas, documentales, paisajes, etcétera; luego crearon un modelo describiendo las características de las escenas que activaban cada uno de los *voxels*, y finalmente les mostraron nuevos videos a los sujetos, es decir, videos que no fueron usados para crear los modelos, y en función de las activaciones de los *voxels*, usaron estos modelos para reproducir lo que estaban viendo a cada instante[89].

Como el resto de la comunidad científica, quedé realmente asombrado por los resultados, pero en mi caso la sorpresa fue mayor. A fin de cuentas, Gallant estaba llevando a cabo, en la vida real, el experimento de aquel científico de *Hasta el fin del mundo*. Los resultados distan mucho de ser perfectos, pero su método reproduce, al menos a grandes rasgos, las formas de las cosas que veían los sujetos y, curiosamente, estas reproducciones tienen mucho en común

con aquellas imágenes borrosas que usara Wenders en su película (como la imagen que ilustra este capítulo).

Escenas de los videos mostradas a los sujetos (arriba) y las reproducciones de estas de acuerdo con la actividad cerebral (abajo). Adaptado de Nishimoto *et al. Current Biology*, 2011.

Hasta aquí mostramos que es posible predecir lo que está viendo una persona a partir de las activaciones de *voxels* medidos con resonancia magnética por imágenes. Estos resultados son espectaculares, pero distan bastante de lo que comúnmente entendemos por "leer la mente". ¿Podremos realmente leer el pensamiento usando esta tecnología? Es decir, ¿podremos no sólo predecir lo que está viendo un sujeto, sino también lo que está pensando, o incluso soñando?

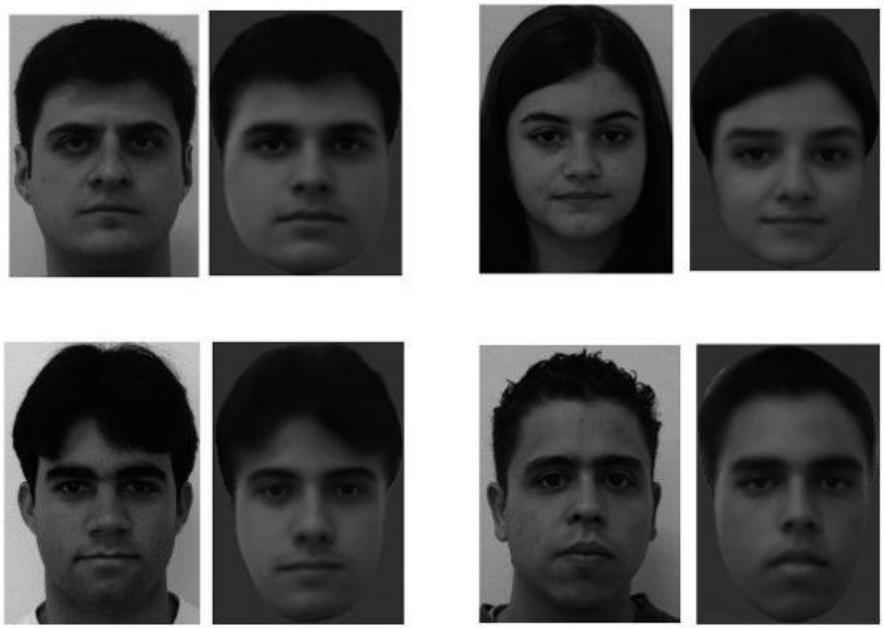
Describimos cómo Adrian Owen llegó a comunicarse con pacientes en estado vegetativo al hacerles imaginar distintas tareas, las cuales producían activaciones análogas a aquellas obtenidas al realizarlas. Otros experimentos también mostraron que se activan las mismas áreas del cerebro tanto al ver como al pensar en caras o paisajes[90]. A partir de estos resultados, parecería entonces que es posible predecir qué está pensando un sujeto espontáneamente, sin necesidad de mostrarle estímulo alguno.

Uno de los aspectos más interesantes de *Hasta el fin del mundo* es la descripción que hace Wenders de la fascinación que sienten las personas al ver imágenes de sus sueños. Pero ¿llegaremos en la vida real a leer los sueños de una persona? Aunque parezca imposible, esto

es justamente lo que hizo Yukiyazu Kamitani y su equipo de investigadores en la Universidad de Kyoto[91]. Primero mostraron a los sujetos un gran número de fotos mientras registraban su actividad cerebral con resonancia magnética por imágenes, para así determinar qué codificaba cada *voxel* y a partir de ello predecir qué veían los sujetos. Hasta aquí nada nuevo. La gran genialidad de Kamitani y su equipo fue que luego pusieron a dormir a los sujetos mientras continuaban registrando su actividad cerebral, para ver si podían predecir sus sueños. Pero ¿cómo verificar si realmente interpretaban sus sueños o cualquier cosa? Muy sencillo: despertaban periódicamente a los sujetos para preguntarles con qué estaban soñando y luego contrastaban sus respuestas con las predicciones que hacían a partir de la actividad cerebral medida poco antes de despertarlos. Los investigadores acertaron el contenido de los sueños en un 60% de los casos, lo cual dista de ser perfecto, pero es una probabilidad muchísimo más alta que aquella obtenida si estuvieran simplemente haciendo predicciones al azar.

Vimos entonces que es posible hacer predicciones sobre lo que está viendo, pensando o incluso soñando una persona usando la actividad cerebral medida con resonancia magnética por imágenes. Vimos también que es posible reproducir las imágenes que está viendo un sujeto a partir de la actividad de los *voxels*. Estas reproducciones son ciertamente limitadas, llegando a discernir solo las formas, los contornos de lo que se muestra y no los rasgos específicos de una cara. Mas esto no deja de ser sorprendente debido a la poca resolución que tiene la resonancia magnética por imágenes: si bien es cierto que obtenemos mucha más información al considerar los *voxels* individualmente y no como parte de una u otra área, un *voxel* representa la actividad conjunta de millones de neuronas y no da información del comportamiento de cada una de estas. Es como tratar de interpretar las circunstancias de lo que está pasando en un estadio escuchando el bramido de la gente a la distancia, sin llegar a discernir lo que dice cada persona. ¿Cuánto más lejos llegaríamos si pudiésemos discriminar la actividad de las neuronas que componen cada *voxel*? Lamentablemente esto ya no es posible con resonancia magnética por

imágenes, pero sí al implantar electrodos en el cerebro.



Caras mostradas al mono (izquierda) y reproducciones de estas hechas a partir de la actividad de un par de cientos de neuronas en los *face patches* (derecha). Adaptado de Chang and Tsao. *Cell*, 2017.

Muchas de las grandes ideas en ciencia son relativamente sencillas. En general, lo que sorprende de estas ideas no es tanto su complejidad intelectual, sino el hecho de que no se le hayan ocurrido antes a nadie. Hace unos años, Doris Tsao y Winrich Freiwald tuvieron una de ellas. Primero hicieron registros con resonancia magnética por imágenes en monos, mientras les mostraban cientos de caras y otros objetos; luego identificaron seis áreas específicas que codificaban preferentemente las caras (que hoy día se conocen como *face patches*), y finalmente introdujeron electrodos para registrar la actividad de neuronas en estas áreas, mostrando que casi todas las neuronas en los *face patches* responden preferentemente a caras[92]. En una serie de experimentos espectaculares, Doris y Winrich han ido dilucidando las funciones específicas de los *face patches* en la codificación de caras[93]. La gran diferencia con los trabajos anteriores es que en estos se introducían electrodos en grandes extensiones de la corteza visual, esperando

encontrar alguna que otra respuesta a caras, mientras que Doris y Winrich sabían exactamente en dónde implantar los electrodos, incrementando notablemente la probabilidad de encontrar respuestas neuronales.

El resultado más notable se produjo a pocos meses de escribir estas líneas. Tras varios años de estudio, Doris dio con el modelo exacto que caracterizaba casi a la perfección las respuestas de estas neuronas. En otras palabras, encontró a qué características de las caras respondía cada neurona, ya sea el ancho de los labios, el largo de la nariz, la separación entre los ojos o características más complejas. A partir de este resultado surgen dos predicciones realmente espectaculares. Primero, una vez establecido el modelo de aquello a lo que responden las neuronas, sería en principio posible predecir la respuesta de estas a cualquier cara; basta con introducir las características de la cara en el modelo, el cual calcularía las respuestas de las neuronas. Segundo, a partir de la respuesta de un grupo de neuronas, podríamos hacer el proceso inverso y reproducir la cara vista, con muchísima más precisión que la obtenida con resonancia magnética por imágenes. Esto es justamente lo que lograron[94].

La similitud entre las caras mostradas y aquellas reproducidas es extraordinaria. A riesgo de ser reiterativo, resalto este resultado porque es realmente increíble: las reproducciones en la figura de arriba fueron obtenidas solo considerando la actividad de unas 200 neuronas, usando un modelo que combinaba en la proporción adecuada las distintas características (largo de nariz, tamaño de la boca, etcétera) codificada por cada una de ellas. Es como si el modelo estuviera armando un identikit a partir del disparo de las neuronas.

En este punto, más de uno quizás piense que estos resultados con monos, como aquellos con resonancia magnética por imágenes o los que describiera con pacientes epilépticos, en los que logramos que proyectaran sus pensamientos en una pantalla, son fascinantes aunque quizás también algo controversiales. Da miedo el pensar que podemos reconstruir lo que está viendo o pensando una persona con solo medir su actividad cerebral. Sin embargo, cabe aclarar que no podemos realizar predicciones sobre lo que piensa una persona sin que esta

acepte voluntariamente colaborar con este tipo de experimentos, para así poder ajustar los algoritmos a usarse, etcétera. Estamos muy lejos de andar por la calle, como Damiel y Casiel, interpretando los pensamientos privados de aquellos transeúntes con quienes nos cruzamos.

Pero más allá de esta salvedad, de alguna manera constantemente leemos a grandes rasgos los pensamientos de los otros, es decir, quizás no sepamos si una persona con la que hablamos está pensando en sus futuras vacaciones en el Caribe o en el partido de fútbol del domingo, pero sí podemos interpretar algunas de sus emociones. Lo hacemos todo el tiempo y tan naturalmente que en general no somos conscientes de ello. No analizamos la actividad de neuronas o de *voxels*, pero sí escrudiñamos pequeños y casi imperceptibles gestos inconscientes de aquellos con quienes interactuamos para descifrar sus estados mentales; si se alegran o no de vernos, si los aburrimos, si están cansados, tristes, alegres o apurados. Las expresiones faciales en general delatan las emociones y, hasta cierto punto, los pensamientos de una persona[95]. Esta es la base de las interacciones sociales, no solo para estimar si a alguien lo alegra o enfada algún comentario que hagamos, sino hasta para algo tan trivial como interpretar los gestos de nuestro interlocutor y así saber cuándo hablar y cuándo escuchar en una conversación mundana.

¿Qué estado de ánimo expresa esta mirada? ¿Será sarcasmo, seriedad, sospecha o desaliento? Adaptado de Baron-Cohen. *The Psychologist*, 2008.

La incapacidad de leer las expresiones faciales de los otros y,

consecuentemente, de poder relacionarse socialmente, es una de las características del autismo. Los autistas suelen evitar mirar a la cara y particularmente a los ojos. Y justamente son los ojos los que dan más información sobre el estado interno de una persona. A un autista le es difícil interpretar el estado de ánimo que expresa la mirada en la figura de abajo, mientras que para el común de la gente esa mirada claramente identifica un estado de tristeza[96].

Leer la mente, así como en *Hasta el fin del mundo* y otras tantas películas de ciencia ficción, ya no es una quimera; es una realidad que ha dado a luz la ciencia en los últimos años, ya sea al interpretar las señales cerebrales obtenidas con resonancia magnética por imágenes o aquellas producidas por neuronas registradas con electrodos intracraneales. Vimos también que de alguna manera leemos la mente de los otros al interpretar gestos inconscientes. Un mentalista o, sin ir más lejos, un jugador profesional de póker, estudia hasta el hartazgo dichos gestos para tratar de prever lo que piensa una persona. Esta discusión se relaciona mucho con aquella de los capítulos anteriores en donde hablamos sobre la teoría de la mente, la capacidad de ponerse en el lugar del otro e interpretar sus pensamientos. Y no es que nos abrume la curiosidad de querer hurgar en los pensamientos personales de los otros, por el contrario, lo que buscamos es tratar de entender qué sienten aquellos con quienes interactuamos, pues esta es la base de la empatía y de nuestras interacciones sociales.

75 Charlie Dunbar Broad; *The mind and its place in nature*; p. 71; 1925.

76 Para más detalles, ver: Frank Jackson. “Epiphenomenal qualia”. *Philosophical Quarterly* 32: 127-136; 1982.

77 Nagel describe esta idea en Thomas Nagel. “What is it like to be a bat?” *The Philosophical Review* 83: 435-450; 1974.

78 Daniel Dennett; *Consciousness explained*; pp. 398-401.

79 Ibídem; pp: 441-448.

80 Para una descripción más detallada de estos algoritmos ver: Rodrigo Quian Quiroga and Stefano Panzeri. “Extracting information from neural populations: Information theory and decoding approaches”. *Nature Reviews Neuroscience* 10:

173-185; 2009.

81 Rodrigo Quian Quiroga, Leila Reddy, Christof Koch and Itzhak Fried. “Decoding visual inputs from multiple neurons in the human temporal lobe”. *Journal of Neurophysiology* 98: 1997-2007; 2007.

82 Para más detalles, ver: Moran Cerf, Nikil Thiruvengadam, Florian Mormann, Alexander Kraskov, Rodrigo Quian Quiroga, Christof Koch and Itzhak Fried. “Online, voluntary control of human temporal lobe neurons”. *Nature* 467: 1104-1108; 2010.

83 Adrian Owen, Martin Coleman, Melanie Boly, Matthew Davis, Steven Laureys and John Pickard. “Detecting Awareness in the Vegetative State”. *Science* 313: 1402; 2006.

84 Martin Monti, Audrey Vanhaudenhuyse, Martin Coleman, Melanie Boly, John Pickard, Luaba Tshibanda, Adrian M. Owen and Steven Laureys. “Willful Modulation of Brain Activity in Disorders of Consciousness”. *The New England Journal of Medicine* 362: 579-589; 2010.

85 Estos resultados están descriptos en: Uri Hasson, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach. “Intersubject synchronization of cortical activity during natural vision”. *Science* 303: 1634-1640; 2004.

86 James Haxby, Ida Gobbini, Maura Furey, Alumit Ishai, Jennifer Schouten and Pietro Pietrini. “Distributed and overlapping representations of faces and objects in ventral temporal cortex”. *Science* 293: 2425-2430; 2001.

87 El método de análisis desarrollado por Haxby y sus colaboradores ha dado lugar a infinidad de trabajos científicos en los últimos años. Para una revisión de estos trabajos ver: John-Dylan Haynes and Geraint Rees. “Decoding mental states from brain activity in humans”. *Nature Reviews Neuroscience* 7: 523-534; 2006.

Para una descripción más detallada de los métodos, ver: John-Dylan Haynes. “A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives”. *Neuron* 87: 257-270; 2015.

88 Ver por ejemplo: Kendrick Kay, Thomas Naselaris, Ryan Prenger and Jack Gallant. “Identifying natural images from human brain activity”. *Nature* 452: 352-355; 2008.

89 Para más detalles, ver: Shinji Nishimoto, An Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu and Jack Gallant. “Recunstructing visual experiences from brain activity evoked by natural movies”. *Current Biology* 21: 1641-1646; 2011.

90 Kathleen O’Craven and Nancy Kanwisher. “Mental imagery of faces and places activates corresponding stimulus-specific brain regions”. *Journal of Cognitive Neuroscience* 12: 1013-1023; 2000.

91 T. Horikawa, M. Tamaki, Y. Miyawaki and Y. Kamitani. “Neural decoding of visual imagery during sleep”. *Science* 340: 639-642; 2013.

92 Doris Tsao, Winrich Freiwald, Roger Tootell and Margaret Livingstone. “A cortical region consisting entirely of face-selective cells”. *Science* 311: 670-674; 2006.

93 Winrich Freiwald and Doris Tsao. “Functional compartmentalization and

viewpoint generalization within the macaque face-processing system". *Science* 330, 845-851; 2010.

94 Steve Chang and Doris Tsao. "The Code for Facial Identity in the Primate Brain". *Cell* 169: 1013-1028; 2017.

Para una descripción sucinta de estos resultados y su implicancia ver: Rodrigo Quian Quiroga. "How do we recognize a face?" *Cell* 169: 975-977; 2017.

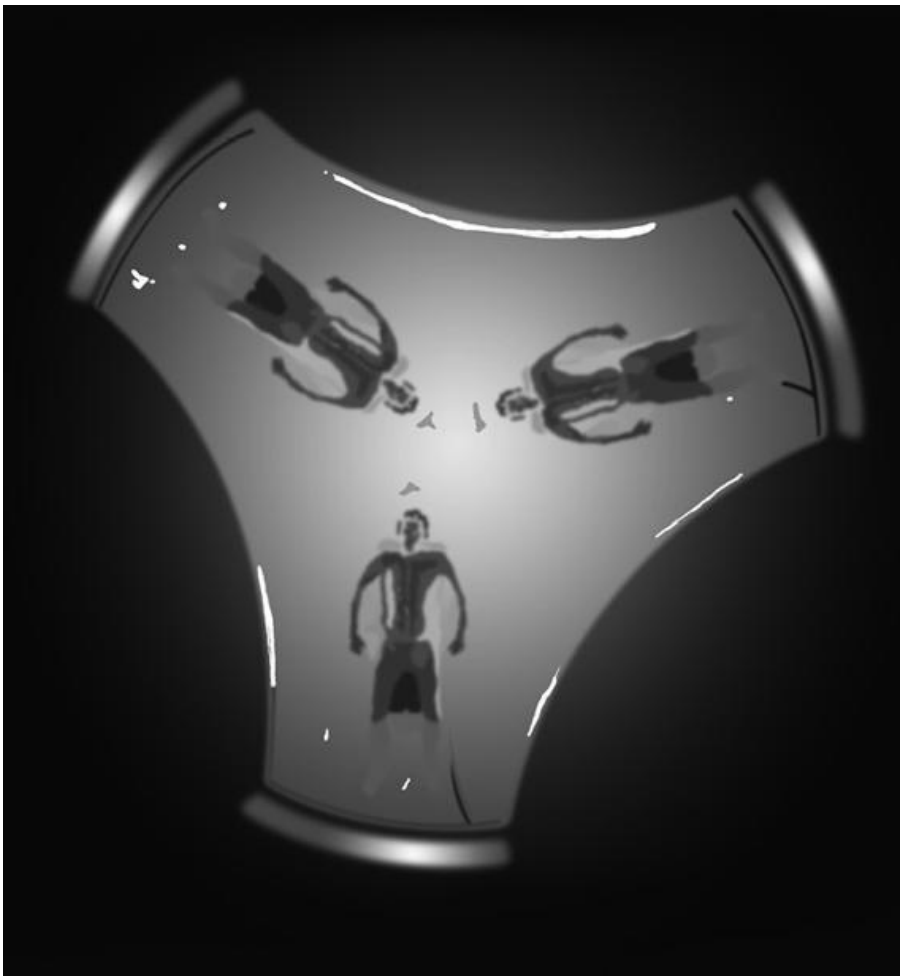
95 El pionero en estudiar cómo las expresiones faciales delatan emociones es el psicólogo norteamericano Paul Ekman, de la Universidad de California, en San Francisco.

96 La interpretación del estado de ánimo que expresan este tipo de miradas es justamente la base de un test para evaluar autismo. Para más detalles, ver: Simon Baron-Cohen, Therese Jolliffe, Catherine Mortimore and Mary Robertson. "Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome". *Journal of Child Psychology and Psychiatry* 38: 813-822; 1997.

Para una descripción más general ver: Simon Baron-Cohen. "Theories of the autistic mind". *The Psychologist* 21: 112-116; 2008.

VI. Minority Report

Libre albedrío



En el capítulo anterior hablamos sobre la posibilidad de leer la mente. Vimos cómo estudios científicos recientes mostraron que efectivamente podemos predecir, basados en la actividad cerebral, lo que está viendo o pensando una persona. Pero ¿podremos también predecir acciones futuras? ¿Podremos, por ejemplo, vaticinar si alguien va a cometer un crimen? De ser así, ¿qué deberíamos hacer en estos casos?

Minority Report fue considerada una de las mejores películas de 2002. Es un *film noir* de ciencia ficción dirigido por Steven Spielberg, basado en un cuento de Philip Dick. La acción transcurre en el año 2054, mayormente en Washington DC, en donde una sección de la policía llamada Precrime logra reducir la tasa de criminalidad literalmente a cero, a partir de predecir —y evitar— crímenes que eventualmente sucederían en un futuro cercano. Las predicciones están basadas en visiones de los *precogs*, tres personas con capacidades psíquicas extraordinarias, mantenidas aisladas, flotando en una gran pileta, concentradas en identificar los distintos crímenes horas antes de que ocurran.

El thriller se dispara cuando uno de los jefes de esta policía, John Anderton, descubre que los precogs predicen que él mismo va a cometer un crimen. Temiendo correr la suerte de los potenciales asesinos que suele arrestar, Anderton escapa y trata de probar un error del sistema. En el fondo, el gran tema de la película es el libre albedrío, la posibilidad de ser realmente libres si nuestro futuro está predeterminado. Anderton tiene la total certeza de ser incapaz de cometer un crimen, aunque, por otro lado, también sabe de la infalibilidad de las predicciones de los precogs y que su suerte está echada.

Como en mi caso, quien haya visto la película muy probablemente no recuerde los detalles de su trama, pero sí algunas escenas

mostrando eventuales avances tecnológicos del futuro; por ejemplo, aquella mostrando a Anderton manejando, cual director de orquesta, las proyecciones de distintos datos e imágenes sobre una pantalla de vidrio con el movimiento de sus manos.

Muchas de las predicciones tecnológicas que se ven en *Minority Report* ya son realidad o no están muy lejos de serlo. De hecho, la manipulación de fotos y archivos con gestos de la mano fue incorporada hace años en los celulares inteligentes y las tablets. Anderton también usa un auto que lo lleva de un punto a otro de la ciudad sin tener que realizar movimiento alguno y, si bien aún no los vemos circular por las calles, los vehículos autónomos ya casi son una realidad, con prototipos de Google y Tesla, entre otros, que en pruebas experimentales llevan recorridos miles de kilómetros. Otra predicción, en este caso estremecedora, es la que muestra a Anderton siendo bombardeado con infinidad de anuncios personalizados al caminar por la calle, sugiriéndole una marca de auto, tomar una cerveza y demás formas de gastar su dinero. Por un lado, el reconocimiento automático de personas partiendo de los rasgos faciales ya es una realidad, y por otro, Internet nos satura constantemente con avisos personalizados; no tenemos más que abrir la página de un diario para ver sugerencias de Amazon según hayan sido nuestras búsquedas previas. Juntando estos dos avances, no es una quimera pensar que sistemas de publicidad puedan reconocer a un transeúnte al pasar y ofrecerle distintos productos de acuerdo con sus gustos. *Minority Report* también muestra otras tecnologías que ya son de uso común y que en aquel momento no existían: relojes y anteojos inteligentes, parecidos a los que se consiguen en la actualidad, una casa inteligente controlada por la voz de su dueño, etcétera.

La cantidad de aciertos en las predicciones de *Minority Report* es realmente notable. Pero no es casual. Antes de filmar la película, en 1999, Spielberg reunió a varios líderes de ciencia y tecnología durante tres días en un hotel en California. El objetivo era interactuar libremente y tratar de vislumbrar cómo sería el futuro en 50 años. Las ideas revolucionarias de la película fueron el resultado de dichas interacciones y, más aún, los *gadgets* mostrados eran tan atractivos que

infinidad de estudiantes, emprendedores y científicos se abocaron a tratar de replicarlos en la vida real.

Ese *feedback* entre visionarios imaginando un futuro posible y emprendedores abocados a volver realidad tales predicciones, hacen de hecho casi imposible vislumbrar cómo viviremos en unos años. No debimos esperar hasta 2054, el año en que trascurre la acción. Muchas de las ideas de *Minority Report* ya son realidad tan solo 15 años después de su estreno.

Pero el tema de este capítulo no es tanto sobre desarrollos tecnológicos, sino sobre la predicción más profunda de la película: ¿podremos predecir el comportamiento de las personas? ¿Podremos juzgar a alguien basándonos en sus actos futuros? Esta es una discusión que ha venido haciendo estragos en la filosofía desde hace siglos, pues toca muy de cerca concepciones sobre la libertad de las personas y la responsabilidad moral que tienen de sus actos.

Isaac Newton, Nicolás Copérnico y Galileo Galilei, tres pilares de la revolución científica del siglo XVII.

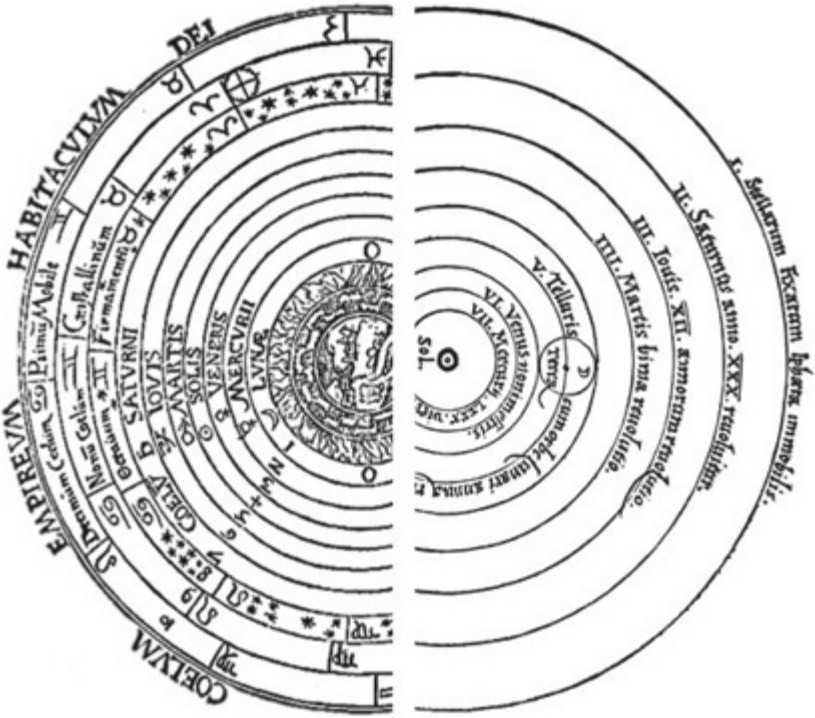
En 1687, Isaac Newton (1643-1727) publica *Principios naturales de la filosofía matemática* (*Philosophiæ Naturalis Principia Mathematica*), en donde sienta las bases de la formulación moderna de la mecánica clásica. La obra de Newton marca el punto más elevado de la revolución científica del siglo XVII, que en la física daría por tierra con el postulado de Aristóteles (384-322 a. C.) sobre el movimiento de los cuerpos y la concepción geocéntrica del universo de Ptolomeo (100-170 d.C.).

Aristóteles sostenía que para mantener a un cuerpo en movimiento

hace falta aplicarle constantemente una fuerza. Esta idea parece muy razonable: mientras el caballo tira del carro lo mantiene en movimiento y, si deja de tirar, el carro se detiene. Un principio distinto regía el movimiento de los cuerpos celestes en el firmamento. Ptolomeo argumentaba que estos realizan órbitas alrededor de la Tierra, lo cual también parecería tener mucho sentido, ya que todos los objetos caen hacia el centro de la Tierra y vemos a los planetas y estrellas girar alrededor de esta. Pero ¿qué empuja a los objetos celestiales en su trayectoria? Aristóteles afirmaba que la Tierra estaba compuesta de cuatro elementos —agua, aire, tierra y fuego—, mientras que en el firmamento existía un quinto elemento, el éter, que estaba en constante movimiento.

Hoy parecería muy fácil refutar los argumentos de Aristóteles y Ptolomeo, pero en realidad no es tan así. Después de todo, tuvieron que pasar más de 1200 años para que surgiera la revolución científica. Por supuesto, en la actualidad nadie duda de que la Tierra no es el centro del universo, pero imaginemos que pudiéramos viajar en el tiempo y discutir mano a mano con Aristóteles y Ptolomeo. ¿Cómo los convenceríamos de que efectivamente es la Tierra la que gira alrededor del Sol y no al revés?

La revolución científica del siglo XVII comenzó a gestarse un par de siglos antes de Newton, cuando Nicolás Copérnico (1473-1543), astrónomo y matemático polaco, propuso un modelo para estimar la trayectoria de los planetas, asumiendo que estos giran alrededor del Sol. El gran problema del modelo de Ptolomeo era el movimiento retrógrado observado en algunos planetas como Marte: día tras día, Marte recorre una trayectoria en el firmamento, pero cada tanto retrocede.



Modelo geocéntrico de Ptolomeo (izquierda) y heliocéntrico de Copérnico (derecha).

La compleja explicación de Ptolomeo era que los planetas giran alrededor de puntos específicos llamados *epiciclos*, los cuales a su vez giran alrededor de la Tierra. El modelo de Copérnico —evitando estos epiciclos, al considerar que los planetas giran alrededor del Sol— era más sencillo, pero aún no brindaba predicciones precisas del movimiento de los planetas. Tales predicciones recién serían posibles luego de que el astrónomo y matemático alemán Johannes Kepler (1571-1630) rompiera con la idea de recorridos a lo largo de círculos perfectos de Aristóteles y postulara una trayectoria elíptica de los planetas alrededor del Sol. La precisión y simpleza de los cálculos de Kepler constituyen en sí mismos un argumento convincente del modelo heliocéntrico, pero la evidencia experimental irrefutable llegaría poco después, a partir de las observaciones de Galileo Galilei (1564-1642).

Tras haber mejorado dramáticamente la resolución del telescopio, Galileo descubrió, entre otras cosas, las lunas de Júpiter y,

consecuentemente, que los cuerpos celestes no necesariamente giran alrededor de la Tierra. Por otro lado, Galileo plantearía el *principio de inercia*. Contrariamente a lo que creía Aristóteles, los cuerpos a los que deja de aplicárseles una fuerza no se detienen, sino que continúan su movimiento. El carro no se detiene porque el caballo deja de tirar, sino por la acción de fuerzas de fricción que lo frenan. Usando ejemplos de nuestros días, si quitamos el pie del acelerador, un auto sigue andando un tiempo hasta que eventualmente se detiene por el rozamiento con el aire y en las ruedas; una nave en el espacio exterior continua con su movimiento eternamente a no ser que alguna fuerza, como la gravedad de algún planeta cercano o la propulsión de sus cohetes, cambie su trayectoria. Por lo tanto, según el principio de inercia de Galileo ya no hace falta imaginar un éter en movimiento para entender el desplazamiento de los planetas; a falta de rozamiento u otra fuerza que los detenga en el espacio exterior, estos continúan su trayectoria (elíptica) atraídos por la fuerza de gravedad del Sol.

Basado en las observaciones de Galileo, Newton luego sintetizaría en tres leyes fundamentales los principios que rigen el movimiento de los cuerpos y propondría la *ley de gravitación*, a partir de la cual deduce las trayectorias elípticas de los planetas postulada por Kepler. El gran avance de Newton fue demostrar que los principios que rigen la mecánica de los cuerpos en la Tierra —la caída de una manzana— son los mismos que rigen los movimientos de los cuerpos celestes: la trayectoria de la Luna alrededor de la Tierra, o de la Tierra alrededor del Sol. A partir de Newton, la física finalmente describía el movimiento de los cuerpos —de todos los cuerpos— a partir de leyes fundamentales sin necesidad de recurrir a argumentaciones extrañas (la existencia de un éter en movimiento, o planetas girando alrededor de epiciclos) para explicar observaciones experimentales.

En este contexto, nacido en los albores de la revolución científica (más de un siglo después de Copérnico y contemporáneo de Kepler y Galileo), René Descartes (1596-1650) comprendió que el avance de la ciencia no debe estar limitado por preconcepciones pasadas, como aquellas de Aristóteles y Ptolomeo, y planteó la duda sistemática como fuente del conocimiento. Dice al respecto (*Discurso del método*, parte

4):

> Así puesto que los sentidos nos engañan […] y puesto que hay hombres que yerran al razonar […] rechacé como falsas todas las razones que anteriormente había tenido por demostrativas […] resolví fingir que todas las cosas que hasta entonces habían entrado en mi espíritu no eran más verdaderas que las ilusiones de mis sueños.

Como viéramos en el capítulo 4, en sus *Meditaciones* Descartes ilustra esta idea concibiendo un genio maligno que, haciendo uso de un poder ilimitado, es capaz de engañarnos sobre la realidad. Tal duda sistemática lo llevaría a encontrar su verdad más fundamental —pienso, luego existo— y a iniciar la revolución, contemporánea a la revolución científica, que daría lugar al nacimiento de la filosofía moderna.

A partir de la publicación de los Principios de Newton, la física tuvo un período de apogeo como nunca antes en su historia. Y así como Descartes imaginara un genio maligno capaz de engañar los sentidos, el matemático francés Pierre Laplace (1749-1827) concebiría, tiempo después, un genio análogo capaz de predecir la posición de cada partícula en el universo usando las leyes de la mecánica de Newton. Dice Laplace (*Essai philosophique sur les probabilités*, p. 4):

> Hay, pues, que considerar el estado actual del universo como efecto de su estado precedente y como causa del que lo sucederá. Una inteligencia que en un determinado instante pudiera conocer todas las fuerzas que impulsan la naturaleza y la respectiva posición de los seres que la componen —una inteligencia lo suficientemente vasta para analizar esos datos— abarcaría en una sola fórmula los movimientos de los mayores cuerpos del universo y aquellos de los más ínfimos átomos; para ella, nada sería incierto y tanto el pasado como el futuro estarían presentes ante sus ojos.

El argumento de Laplace es lo que se conoce como *determinismo*: cada estado del universo es la consecuencia inevitable y única de estados anteriores. En otras palabras, conociendo las leyes del universo, el estado inicial de un objeto determina sus estados futuros.

Y esto no solo se aplica a manzanas y planetas en movimiento, sino también a nosotros mismos.

Nuestro cerebro no es más que la actividad de miles de millones de neuronas. Entonces, si el genio de Laplace conociera el estado de nuestras neuronas en un momento dado, podría también predecir sus estados futuros. Por supuesto, esta tarea va mucho más allá de nuestra capacidad de medición y de cálculo. Es imposible medir la actividad de todas nuestras neuronas (y de cada partícula que compone nuestro cuerpo, ya que el cerebro no está aislado del cuerpo), y aunque tuviéramos tal información sería también imposible calcular estados futuros. Pero dejemos de lado limitaciones prácticas y sigamos con el argumento teórico.

En principio, el genio de Laplace conocería el estado de mis neuronas desde mi nacimiento hasta el día de mi muerte. Y al conocer el estado de mis neuronas (y de mi cuerpo y de todo lo que interactúa conmigo), podría predecir si dentro de unos segundos se activarán neuronas en mi corteza motora que me harán levantar un brazo, o aquellas que dan lugar al movimiento preciso de mis dedos en el teclado, determinando las palabras que escribo en este momento. La conclusión es devastadora: según el determinismo, todo lo que hago está predeterminado, es inevitable; es la consecuencia lógica de la aplicación de las leyes de la física, las cuales no dan lugar a posibles alternativas. Y este es el mazazo tremendo que la física le propina a la filosofía: si todo está predeterminado, ¿somos libres?, ¿podemos elegir nuestras acciones futuras? En el fondo nos estamos preguntando si realmente existe el libre albedrío[97].

Los debates sobre el libre albedrío son de larga data en la filosofía y arrancan mucho antes de la revolución científica. Como viéramos antes, en la antigua Grecia Demócrito pregonaba que en el universo no hay más que átomos y el vacío entre estos. En este mundo material nada ocurre al azar; todo tiene una razón de ser. Con estos argumentos Demócrito buscaba dejar de lado la idea de que nuestros actos están determinados por la voluntad de los dioses de la Antigüedad, afirmando que tienen causas terrenales. Sin embargo, a pesar de dejar de lado una voluntad sobrenatural que explique nuestro

comportamiento, Demócrito termina pregonando el determinismo que luego florecería en la revolución científica y que terminaría desafiando la idea del libre albedrío.

Santo Tomás de Aquino (1225-1274), el más notable de los filósofos escolásticos de la Edad Media, y anteriormente san Agustín (354-430), vieron la necesidad de aludir a uno de los grandes conflictos filosóficos de la religión, el cual está justamente ligado a la noción del libre albedrío : si existe un Dios que conoce de antemano cada uno de nuestros actos, ¿somos entonces libres?, ¿podríamos optar por actuar de otra manera? Más aún, si hay un Dios bueno y todopoderoso, ¿por qué existe el mal? ¿Cómo puede un buen cristiano ser libre, si no puede optar por el pecado?

La respuesta de santo Tomás y de san Agustín son la base de la actual doctrina de la iglesia. Por un lado, las decisiones las tomamos libremente. Dios conoce el resultado de nuestras elecciones, pero no nos fuerza a actuar de una u otra manera. El pecado es para ellos un abuso del libre albedrío (ya presente en el pecado original de Adán y Eva). El buen cristiano debe siempre hacer el bien —si no es esclavo del pecado— pero tiene libertad de elegir la mejor manera de hacerlo.

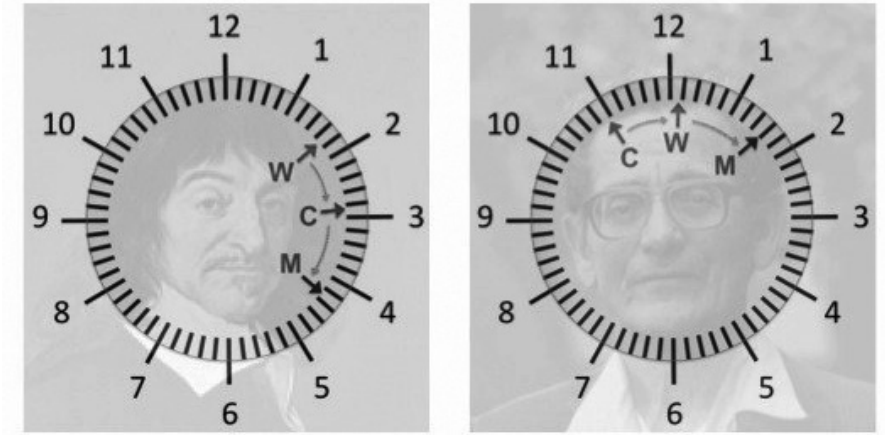
Descartes entendía al libre albedrío como la capacidad de elegir hacer o no hacer una cosa. Dada la separación que hiciera entre mente y materia, para Descartes determinismo y libre albedrío son compatibles, lo cual en filosofía se conoce como *compatibilismo*. Según esta visión, el determinismo viene dado por las leyes de la naturaleza y rige el comportamiento de los cuerpos y los animales. En el dominio de lo material, cada evento está unívocamente predeterminado por eventos anteriores. Sin embargo, el hombre también posee una mente con la cual hace uso de su voluntad y elige libremente su futuro. El libre albedrío ejercitado por la mente es el origen de los eventos físicos que se desencadenan en el reino de la materia.

Existe entonces el determinismo, pero el hombre puede cambiar la secuencia de estos eventos a partir de tomar libremente decisiones con su mente. Por ejemplo, si levanto un brazo, es mi mente la que decide realizar esta acción y luego (según Descartes a través de la glándula pineal) transmite esta orden al cerebro y a los distintos nervios y

músculos que llevarán a cabo el movimiento. La decisión consciente de realizar una acción precede su ejecución. Nada parecería tener más sentido. Sin embargo, la neurociencia de los últimos años ha derrumbado por completo el argumento de Descartes. Ya vimos que son muy pocos los que hoy día creen en una separación entre mente y cerebro, pero la refutación de la noción de libre albedrío de Descartes se daría a partir de una serie de experimentos con resultados demoledores.

En 1983, el fisiólogo norteamericano Benjamin Libet reportó los resultados del que quizás sea uno de los experimentos más controversiales en la historia de la neurociencia[98]. La idea del experimento es muy sencilla: los participantes miraban una pantalla en donde un punto recorría una trayectoria circular (como el segundero de un reloj) y, en el momento de su elección, no tenían más que flexionar los dedos y luego reportar la posición del segundero en el momento en que creían haber tomado la decisión consciente de hacer el movimiento. En simultaneo, Libet registró la actividad electromiográfica (EMG) —la actividad muscular medida con electrodos en el antebrazo— para determinar el momento preciso del inicio del movimiento, y también registró la actividad electroencefalográfica (EEG), para identificar el momento de activación del cerebro precediendo el movimiento. (Desde hacía años se conocía que la ejecución de un movimiento es precedida por una activación en el EEG llamada *Bereitschaftspotential* o *readiness potential.*)

En conclusión, Libet podía medir tres tiempos a partir de este experimento: 1) el tiempo de inicio del movimiento, a partir del EMG; 2) el tiempo en que el sujeto reportaba ser consciente de iniciar el movimiento, al cual llamó “momento W” (la primera letra de la palabra inglesa *will*), y 3) el tiempo de preparación del movimiento en el cerebro, a partir de la señal EEG.



Según Descartes (izquierda), la elección consciente de iniciar un movimiento (W) precede a la activación en el cerebro (C), que a su vez precede el movimiento muscular (M). Sin embargo, Libet (derecha) demostró que la activación cerebral precede a la elección consciente, poniendo en jaque la idea de libre albedrío.

A partir de estos experimentos, Libet encontró que el momento de elección consciente (W) se daba alrededor de 200 milisegundos antes del inicio de la acción motora (M). Sin embargo, lo sorprendente fue que la activación cerebral (C) no solo ocurría antes que la activación motora (M), sino que también precedía 300 milisegundos al momento de la elección consciente (W). Este hallazgo dio por tierra con la idea de libre albedrío de Descartes. La activación cerebral no es entonces una consecuencia de la elección consciente de la mente, sino que la precede. En otras palabras, la decisión consciente es una consecuencia de procesos inconscientes en el cerebro (que luego dan lugar a las activaciones motoras). Pero si nuestras acciones se deben a la activación inconsciente de neuronas —de hecho, siguiendo reglas deterministas— aparentemente no queda más remedio que aceptar que el libre albedrío no existe.

La mayoría de los científicos contemporáneos cree que el libre albedrío es una ilusión o, al menos, que no es lo que habitualmente se entiende por ello. Esta visión ya fue pregonada por Baruch Spinoza (1632-1677), quien junto a Leibniz y Descartes fuera uno de los filósofos racionalistas del siglo XVII. En un pasaje en donde, con

argumentos muy sólidos y en línea con la visión de los científicos actuales, destroza la noción de dualismo cartesiano y la separación entre mente y materia, dice Spinoza sobre el libre albedrío (Ética. Parte III, Proposición II):

> La experiencia misma, no menos claramente que la razón, enseña que los hombres creen ser libres solo porque son conscientes de sus acciones e ignorantes de las causas que las determinan […]. Así pues, quienes creen que hablan, o callan, o hacen cualquier cosa por libre decisión de su mente, no hacen más que soñar con los ojos abiertos.

El experimento de Libet no solo abrió grandes controversias dentro de la filosofía, sino que también fue muy criticado en neurociencia. El principal punto débil es que la conclusión desoladora de que el libre albedrío no existe se basa en la estimación del punto W, el momento exacto en que los sujetos alegan ser conscientes de decidir realizar una acción[99]. Esto es obviamente poco preciso y subjetivo. Si me paro frente a un reloj esperando decidir levantar un brazo, me es muy difícil determinar el momento exacto en el que tomo esta decisión y un error de tan solo una fracción de segundo podría ser el origen de creer que la activación cerebral precede la decisión consciente. Sin embargo, experimentos más recientes confirman que la conclusión de Libet era correcta.

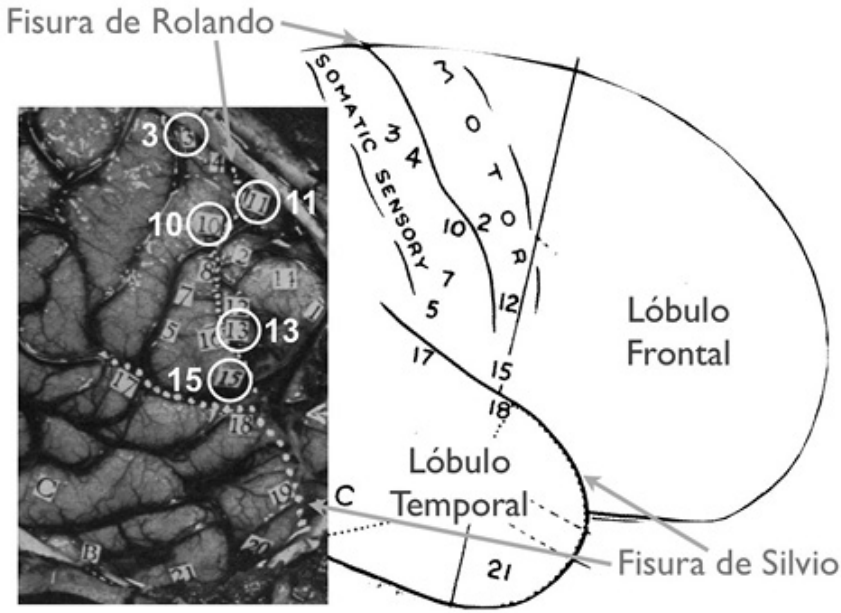
John-Dylan Haynes, un neurocientífico formado en Londres y actualmente profesor en Berlín, hizo su carrera usando las técnicas descriptas en el capítulo anterior, analizando patrones de activaciones de *voxels* medidos con resonancia magnética por imágenes, para predecir lo que están viendo o pensando los sujetos. Haynes decidió reflotar la idea del experimento de Libet —prediciendo la activación cerebral que eventualmente da lugar a un movimiento— pero usando las activaciones de los *voxels* en vez de señales EEG. La ventaja es que, como vimos anteriormente, las respuestas observadas en el EEG vienen dadas por la aparición del *readiness potential*, el cual al ser generado en la corteza motora suplementaria refleja activaciones relativamente tardías, ya que esta área es una de las últimas

involucradas en la generación del movimiento. Pues entonces, probablemente activaciones anteriores que no son observables con EEG también reflejen la decisión de efectuar un movimiento, y mucho antes que los 500 milisegundos reportados por Libet.

El resultado sorprendente de Haynes fue que activaciones vistas en la resonancia magnética por imágenes podían predecir el movimiento de los sujetos hasta 10 segundos antes de que lo hicieran. Por lo tanto, si bien es cierto que es difícil determinar con precisión el momento exacto en el que los sujetos toman la decisión consciente de efectuar un movimiento (lo cual lleva a dudar de los resultados de Libet), en el caso de Haynes no queda duda de que activaciones cerebrales inconscientes preceden por varios segundos la decisión de realizar dichas acciones.

Diversos resultados experimentales ofrecen aún más evidencia en contra de la concepción que naturalmente tenemos sobre el libre albedrío. En el capítulo 4 hablamos de Wilder Penfield y su protocolo de estimulación eléctrica en pacientes epilépticos con el cerebro expuesto. El objetivo de estas estimulaciones era localizar distintas funciones cerebrales y evaluar la extensión de una potencial resección quirúrgica del foco epiléptico que no afecte dichas funciones. Tras estudiar miles de pacientes con este método, Penfield hizo un mapa de la corteza sensorial y motora en el cerebro humano, localizando áreas cuya activación generaba, por ejemplo, movimientos o cosquilleos en distintas partes del cuerpo.

Al hablar de Penfield también nos referimos a los experimentos de mi colega Itzhak Fried, quien al estimular un área del cerebro evocaba la risa del paciente. Más aún, en otro grupo de pacientes Ithzak mostró que estimulando podía no solo generar, como Penfield, movimientos específicos, sino también la predisposición de hacerlos[100]. Los resultados de Penfield y Fried muestran entonces que, lejos de tener una mente que origina las decisiones conscientes, la determinación de realizar un movimiento, o hasta el impulso mismo de llevarlo a cabo, puede ser generado a partir de estimular eléctricamente neuronas en distintas áreas del cerebro.



Puntos estimulados por Penfield en un paciente (izquierda) y croquis de la ubicación de dichos puntos en el cerebro (no todos los puntos aparecen en el croquis). La fisura de Rolando separa la corteza sensorial de la motora y la fisura de Silvio separa el lóbulo frontal del temporal. Al estimular en los puntos 3, 10 y 15, en la corteza sensorial, el paciente reportó tener un cosquilleo en el pulgar izquierdo, un sensación en el labio inferior izquierdo y un adormecimiento de la lengua, respectivamente. Al estimular en los puntos 13 y 11, en la corteza motora, Penfiel generó un movimiento en el lado izquierdo de la cara (punto 13) y que el paciente abriera la boca pronunciando repetidamente la vocal “A” (punto 11)[101].

Para estimular el cerebro no necesariamente hace falta realizar una cirugía. También puede hacérselo con métodos no invasivos, a través de pulsos magnéticos aplicados sobre el cuero cabelludo. Esto es lo que se conoce como *estimulación transcranial magnética*. Uno de los expertos en su uso es Álvaro Pascual-Leone, un científico español que trabaja desde hace varios años en la Universidad de Harvard, en los Estados Unidos. Según dónde es aplicada, la estimulación transcranial magnética genera efectos sorprendentes, como arrestos del habla, cosquilleos, movimientos involuntarios, etcétera. En un estudio notable, Pascual-Leone y sus colegas pidieron a un grupo de sujetos

que extendieran el dedo índice, ya sea de la mano derecha o de la izquierda, en el momento de su elección[102]. Hasta aquí, esto no es más que un refinamiento del experimento de Libet: los sujetos no solo elegían cuándo extender un dedo, sino también qué dedo iban a extender. Pero lo interesante es que cada tanto los investigadores aplicaban estimulación transcranial magnética, provocando la extensión de uno u otro dedo, según dónde estimulaban. Usando esta técnica, efectivamente lograron manipular las decisiones de movimiento de los sujetos, pero sorprendentemente estos no eran conscientes de tales manipulaciones, pues reportaban tomar las decisiones libremente, sin notar que en realidad el movimiento no era más que la consecuencia mecánica de la estimulación aplicada.

A partir de estos estudios, queda claro que la aparente libertad que tenemos de tomar decisiones puede ser manipulada sin siquiera ser conscientes de ello. Pero la ilusión de libre albedrío es crucial, pues por más ficticia y maleable que sea, la carencia de esa reflexión consciente que nos da la sensación de estar en control de nuestros actos genera patologías devastadoras.

En “Cuello de gatito negro” (*Octaedro*, 1974), Julio Cortázar relata un encuentro atípico en un vagón del metro en París. Lucho, un argentino enamoradizo, es abordado por una mulata que, discretamente al principio y luego mucho más abiertamente, comienza a acariciarle las manos, sostenidas en la barra del metro. Lo curioso es que el accionar de las manos de la mulata, envueltas en guantes negros, actuaban *motu proprio*, sin la aprobación consciente de su dueña, que infructuosamente trataba de disculparse ante Lucho. Tal comportamiento termina desencadenando una situación bizarra y, por más disparatado que parezca, conductas similares han sido descriptas en pacientes que alegan no tener control de sus manos. Esto es lo que se conoce como *síndrome de la mano ajena*.

Una caracterización desopilante de esta condición fue hecha por Peter Sellers en *Dr. Strangelove*, interpretando a un alocado científico en silla de ruedas que no puede controlar las acciones de su brazo derecho. En uno de los puntos más hilarantes de la película, el brazo rebelde es retenido y castigado por el otro brazo, pero finalmente se

libera para realizar un saludo nazi, darle un golpe de puño a *Dr. Strangelove* y luego tratar de ahorcarlo[103].

El síndrome de la mano ajena viene generalmente dado por una lesión en el cuerpo calloso —el grupo de nervios que conecta ambos hemisferios cerebrales— y la corteza frontal. Un paciente, por ejemplo, era despertado por su mano izquierda, que trataba de ahorcarlo mientras dormía y durante el día le desabotonaba la ropa o luchaba con su otra mano cuando intentaba contestar el teléfono. Parece irreal, de hecho, el hombre decía que era como si "alguien de la luna estuviera controlando su mano". En otro caso, la mano rebelde tomó un lápiz y comenzó a hacer garabatos; más cuando se le hizo notar lo que estaba haciendo, el paciente se sorprendió y reprimió inmediatamente el miembro rebelde con su otra mano. Según el paciente, era la mano y no él quien hacia estas cosas[104].

Sobre la base de los experimentos antes descriptos, es la actividad inconsciente del cerebro la que inicia una acción; poco tiempo después somos conscientes de la acción que vamos a ejecutar e, ignorantes de los procesos inconscientes que la preceden, atribuimos al proceso de pensar en tal acción la intención (o causa) de hacerla. Esto es justamente lo que genera la ilusión del libre albedrío —el creer que es la intención consciente la que causa nuestros actos— y, como muestran los casos de síndrome de la mano ajena, es realmente perturbador cuando esta ilusión no funciona y los procesos inconscientes directamente dictan el funcionamiento de una mano que en apariencia no responde a nuestra voluntad[105].

La sensación de falta de control es todavía mucho más dramática en pacientes con esquizofrenia. Estos pacientes no solo sienten no tener control sobre sus actos, sino que a veces tampoco pueden controlar sus emociones o pensamientos. Es como si estuvieran manejados por agentes externos.

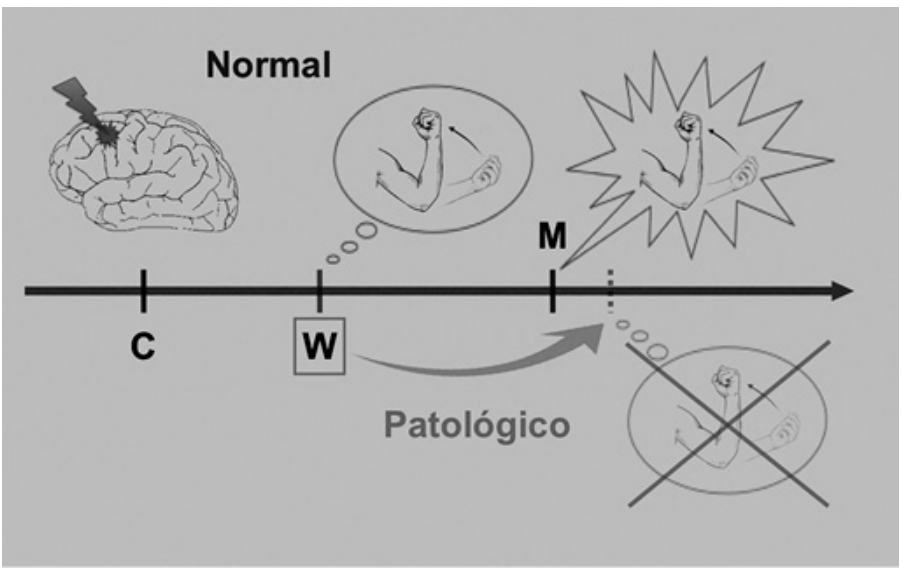
Una caracterización muy popular de esta patología es la de Smeagol (o Gollum) en *El señor de los anillos*, o la de Russell Crowe en su interpretación del famoso economista John Nash en *Una mente brillante*. Los pensamientos de Smeagol o Nash son cooptados por sujetos inexistentes y, en algunos casos, los pacientes con

esquizofrenia incluso alegan que su mente está dominada por extraterrestres (después de todo, ¿quién puede tener semejante poder como para dominar los pensamientos de otra persona a la distancia?). Sin embargo, la neurociencia ofrece una explicación mucho más sencilla[106]: probablemente lo que falle es la reflexión consciente que estas personas tienen sobre sus propias acciones; la ilusión del libre albedrío. Es decir, si falla el proceso de ser conscientes de las acciones que vamos a llevar a cabo, entonces estas se vuelven inesperadas y aparentemente responden a agentes externos. Pero, por supuesto, tales acciones —ya sea el movimiento de un brazo o las palabras que hilamos en nuestro pensamiento— no son generadas por otros, sino por procesos inconscientes en nuestro cerebro. De hecho, es común que pacientes con esquizofrenia reporten escuchar voces extrañas, pero esto no es más que su propia voz, la cual no pueden predecir al no ser conscientes de lo que van a decir antes de hacerlo[107]. Pongámonos por un momento en su lugar e imaginemos hablar, pero sin tener idea de lo que decimos, sorprendiéndonos de las palabras que salen de nuestra boca. En tal situación, no sería raro imaginar que alguien ha tomado el control de nuestro pensamiento.

Uno de los motivos por el cual falla la ilusión del libre albedrío es por retardos en la percepción consciente que tenemos de las cosas. Si la aparente intención de realizar un movimiento (el momento W de Libet), no precede su ejecución, la ilusión del libre albedrío se diluye: al notar hacer un movimiento antes de tener la sensación de haberlo planeado, pareciera que el movimiento no responde a nuestra voluntad[108].

Una demostración muy elegante de cómo retardos temporales influyen en la percepción que tenemos de originar nuestros actos fue dada a través de un experimento en el cual los sujetos operaban un robot que les hacía cosquillas con una varilla de punta suave en la palma de la mano[109]. Por más que tratemos, no podemos hacernos cosquillas a nosotros mismos; es imposible, no hay manera. Pero si es otro quien dirige el movimiento, aunque en el fondo realice el mismo tipo de acciones que las nuestras, saltamos hasta el techo. La gran diferencia es que al ser otro quien realiza los movimientos no tenemos

control, no podemos predecir lo que va a pasar. Lo interesante del experimento en cuestión es que la varilla que hacía cosquillas a los sujetos en la palma de su mano izquierda reproducía los movimientos que estos realizaban con una palanca que movían con su mano derecha. Cuando la réplica del movimiento era instantánea, la sensación de cosquillas era mínima, pero a medida que se introducía un mínimo retardo en la reproducción del movimiento, aunque solo fuera de una décima de segundo (que los sujetos no llegaban a notar), aparecía la sensación de cosquillas. Esto muestra que retardos en la percepción de una acción llevan a creer que la acción la realiza otro.



Activaciones inconscientes en el cerebro (C) preceden la reflexión consciente de, por ejemplo, el acto de flexionar un brazo (W), que a su vez precede la ejecución de la acción motora (M). Retardos en el procesamiento de la reflexión consciente (abajo) hacen que la ilusión de controlar los movimientos (o el habla o los pensamientos) se diluya, dando lugar a comportamientos patológicos, como en el caso de pacientes con esquizofrenia.

Según los experimentos de Penfield, Fried and Pascual-Leone vimos que las decisiones pueden ser manipuladas a partir de estimular eléctrica o magnéticamente grupos de neuronas, sin que los sujetos sean conscientes de las manipulaciones llevadas a cabo. Las estimulaciones externas dan lugar a reacciones que podría decirse

fueron “implantadas” en los sujetos. Mas a veces estas reacciones involuntarias son provocadas por mecanismos internos y patológicos, los cuales son igualmente incontrolables.

Cuando trabajaba como investigador en el Caltech conocí a un estudiante con síndrome de Tourette[110]. Estas personas tienen reflejos involuntarios e incontrolables que van desde algún tic hasta proferir insultos, realizar movimientos espasmódicos (o escupir en las situaciones más embarazosas, como en el caso de ese estudiante). A un amigo en común, una vez le escupió el vidrio del auto y en un acto reflejo tiró un manotazo como para limpiar su saliva, para terminar rompiendo la ventanilla. Por supuesto que a mi amigo no le hizo ninguna gracia la rotura del vidrio, pero ¿qué podía decirle? ¿Qué culpa tenía si no podía controlar sus actos?

En el derecho romano un delito se castiga si existe tanto un *actus reus*, es decir, el acto delictivo en sí mismo, y un *mens rea*, es decir, la mente culpable o, en otras palabras, la voluntad de llevar a cabo el delito. En el caso del estudiante con síndrome de Tourette, uno no puede culparlo de sus reacciones involuntarias porque no hay un *mens rea*; él no tiene control sobre sus actos y no puede evitar escupir, insultar o romper la ventanilla de un manotazo. Asimismo, no puede culparse a una persona con síndrome de la mano ajena por las acciones incontrolables de su mano.

Los casos de esquizofrenia son mucho más delicados, ya que alguien con esta patología puede, por ejemplo, cometer un crimen aparentemente forzado por un agente externo que tomó control de sus actos. De hecho, si se demuestra que un criminal tiene esquizofrenia, muy probablemente no termine en una cárcel, sino en un instituto psiquiátrico. El derecho romano en principio tampoco castiga a quien comete un delito estando sonámbulo o bajo la influencia de otra persona, por ejemplo, a través de una sugestión hipnótica (como en la película *The manchurian candidate*).

Todos estos casos carecen de *mens rea*, la intención, el proceso consciente de querer llevar a cabo el delito. Estas personas no ejercen su voluntad, no hacen uso de su libre albedrío. Pero hete aquí el dilema: ¿cómo podemos siquiera pensar en juzgar a alguien, si

argumentamos que el libre albedrío no existe? En otras palabras, vimos que las decisiones conscientes son una consecuencia de procesos inconscientes en el cerebro que siguen las leyes del determinismo y que son, por ende, inevitables.

Kant argumentaba que sin libre albedrío no puede haber moral. Siglos después, un estudio mostró cómo las concepciones sobre el libre albedrío afectan la moral. A un grupo de sujetos se les dio a leer un texto sobre el determinismo y la inexistencia del libre albedrío, mientras que a otro grupo se les asignó un texto neutro. Ambos grupos realizaron un examen y los investigadores comprobaron que el primer grupo mostraba una tendencia más alta a hacer trampa. Los investigadores concluyeron que, al dudar del libre albedrío, los sujetos tendían a socavar el sentimiento que tienen de sentirse una persona, lo cual probablemente les haya dado la excusa para comportarse como quisieran[111].

En una conferencia en Porto tuve el gusto enorme de conocer a Jim Fallon, un experto en psicópatas asesinos. Y no me refiero a personas que por venganza o en un ataque de ira puedan haber matado, incluso de la manera más brutal y macabra. Me refiero a aquellos que como Hannibal Lecter, en la caracterización de Anthony Hopkins, matan a sangre fría, casi como naturalmente, sin el menor dejo de emoción o remordimiento.

¿Qué es lo que falla en estas personas? Tras estudiar decenas de casos, Jim encontró un patrón común en su cerebro, dado por una deficiencia en lo que se conoce como el *sistema límbico*, el sistema que regula las emociones. Después de todo, tiene sentido: la falla en el sistema emocional subyuga el rechazo moral que naturalmente tenemos al siquiera considerar la posibilidad de matar a una persona. La alteración en el sistema límbico era tan notable, que Jim podía distinguir sin lugar a dudas la imagen del cerebro de un psicópata de la de otro sujeto. Pero un buen día, al analizar una imagen de su propio cerebro, encontró que él mismo tenía la misma alteración[112]. Profesor en la Universidad de California en Irvine, padre de familia y casado hacía más de 40 años, Jim sabía que no era un psicópata como las personas que investigaba, pero ¿dónde estaba la diferencia?

Jim argumenta que los psicópatas asesinos no solo tienen una malformación en el sistema límbico, sino que también tienen una historia de abuso en su infancia. Este es el cóctel explosivo. Una persona como Jim quizás podría matar a alguien sin el menor sentimiento de culpa, pero nunca se le ocurriría hacerlo. Por otro lado, una persona con una infancia traumática podría acumular un nivel de agresión tan grande que lo llevaría a considerar un asesinato, pero desistiría de hacerlo a partir de la represión dada por el sistema límbico; la empatía, la conexión emocional con la otra persona. En cambio, alguien que tenga tanto una deficiencia en el sistema límbico como a una infancia traumática, podría tender a buscar resarcimiento y ningún mecanismo interno lo disuadiría de cometer un crimen.

Volvemos entonces al argumento central de *Minority Report*: si evaluaciones como las de Jim Fallon determinan el comportamiento psicópata de una persona, ¿tenemos como sociedad que tomar medidas preventivas para evitar una masacre? ¿Podemos condenar a alguien sin que haya *actus reus*, es decir, antes de que cometa un delito? Para contestar esta pregunta tenemos que considerar que por más fuerte que sea la tendencia fisiológica y social que tenga una persona de ser psicópata, aún tiene capacidad de elección. Amén de que sea cierto, como argumenta Fallon, que todos los psicópatas tienen una alteración en su cerebro y una infancia traumática, no todas las personas con estos dos componentes necesariamente terminan siendo psicópatas. Quizás haya millones con estas características de quienes no sabemos nada porque tienen un comportamiento normal. En otras palabras, las dos condiciones descriptas por Fallon son una condición necesaria pero no suficiente. Michael Gazzaniga, uno de los pioneros en neurociencia cognitiva, lo pone muy claro[113]: no existe alteración alguna del cerebro que lleve a cometer un delito. Una persona podrá tener más o menos mecanismos de contención, pero siempre tiene la capacidad de decisión, es libre de elegir. Sin embargo, pareciera que nos contradecimos, pues ¿cómo podemos hablar de libertad de elegir si ya establecimos que el libre albedrío no existe?

Muchas de las confusiones sobre el libre albedrío vienen dadas por

lo que se entiende al respecto. Hay que distinguir el problema filosófico y neurocientífico del problema moral y del derecho. Es cierto que nuestro comportamiento está dado por el disparo de neuronas que siguen reglas deterministas, pero a nadie se le ocurriría argumentar ante un juez que, dadas las condiciones del universo y el disparo de billones de neuronas, lo que hizo un asesino era inevitable.

En derecho, por libre albedrío se entiende que alguien actúa *motu proprio*, sin ser coaccionado a cometer un delito; una persona no es culpable de, por ejemplo, haber firmado un cheque falso, si demuestra que otra la obligó a hacerlo con un arma en la cabeza. El planteo que venimos haciendo sobre el libre albedrío en neurociencia —de saber que nuestro comportamiento está dado por la actividad de neuronas siguiendo reglas deterministas— va más allá de consideraciones sobre el sistema legal y da lugar a una discusión mucho más profunda y filosófica. Pero dejemos esto para el final del capítulo y sigamos con un tema mucho más ríspido.

Acordamos que no se puede juzgar a una persona por su predisposición a cometer un delito. Pero mucho más delicada es la situación de aquellos que ya han cometido un delito y se discute si pueden o no ser reinsertados en la sociedad a través de una libertad condicional o incluso tras haber cumplido su condena. Esta discusión involucra dos visiones distintas sobre la justicia: retributiva y utilitaria. La *justicia retributiva* enfatiza el castigo justo —lo que merece una persona por el crimen cometido. La *justicia utilitaria* prioriza, en cambio, el beneficio de la sociedad. En principio ninguna de las dos visiones es más severa que la otra. Para la justicia retributiva, el condenado debe cumplir hasta el último día de su condena, ni uno más, ni uno menos. Para la justicia utilitaria, la condena es más flexible porque el condenado puede ser liberado antes si ya no es un riesgo para la sociedad, pero puede también seguir detenido más allá de su condena si aún lo sigue siendo. Esto último es un arma de doble filo y da lugar a controversiales estimaciones sobre el riesgo de reincidencia y la implementación de políticas preventivas.

Solo en Estados Unidos hay más de dos millones de personas en la cárcel. El costo de mantener tantos presos es altísimo y hace unos años

se planteó la posibilidad de liberar a aquellos que hayan cometido delitos menores y que no necesiten estar encarcelados; por ejemplo, conmutándoles las penas por trabajos comunitarios. En síntesis, factores sociales y económicos llevaron a implementar una justicia utilitaria para delitos menores. Pero ¿cómo determinar quién puede ser liberado y quién no? Para afrontar este problema comenzaron a desarrollarse algoritmos que establecen un factor de riesgo de reincidencia de un sujeto, evaluando su prontuario y las respuestas a un cuestionario de más de 100 preguntas. Mas lo que empezó como una decisión aparentemente razonable (después de todo, por qué gastar tanto dinero en presos que no hace falta tener en la cárcel y no destinar esos recursos, por ejemplo, al sistema de salud), se ha convertido en un monstruo que tiene en jaque al sistema legal norteamericano.

El problema es que ese tipo de algoritmos ya se usa en varios estados, y no solo para conmutar condenas en delitos menores, sino también para sugerir decisiones en todo tipo de delitos y durante todo el proceso legal: si una persona debe ser encarcelada o liberada, el monto de una fianza, si debe otorgarse una libertad condicional, o hasta qué tan severa debe ser una condena.

El algoritmo más usado, llamado COMPAS, estima un factor de riesgo que establece la probabilidad de que alguien vuelva a cometer un delito, y esta información es dada a los jueces antes de dictar sentencia. El primer gran inconveniente es que el algoritmo implícitamente discrimina: un joven de 20 años, soltero y sin empleo va a tener un factor de riesgo mucho más alto que una mujer de mediana edad, casada y con hijos. O sea, por exactamente el mismo delito, el joven tendrá un factor de riesgo más alto y eventualmente una condena más elevada. Esto evidentemente viola la noción de justicia retributiva: otorgar la condena justa basándose solo en el delito cometido. Pero las discriminaciones son aún mucho más aberrantes, pues entre los elementos que se usan para estimar el factor de riesgo, la situación social juega un papel preponderante. Una persona de un barrio humilde y con un entorno social precario va a tener un factor de riesgo mucho más alto que otra con el mismo

prontuario pero de un barrio de clase alta; una persona con familiares que hayan cometido delitos va a tener un factor de riesgo más alto que otra que no los tenga. O sea, al menos en parte se juzga en función de las acciones de otros, lo cual es muy explícito en alguna de las preguntas del cuestionario: ¿estuvieron tus padres alguna vez en prisión? ¿Cuántos de tus amigos o conocidos toman drogas ilegales?, etcétera.

Los ejemplos abundan, un chico negro que solo había cometido un robo menor, fue considerado dos veces más peligroso que un blanco de 50 años, con un historial de violencia doméstica, robos mayores y tráfico de drogas. Otro joven negro que resistió un arresto sin ejercer violencia, fue considerado tres veces más peligroso que un blanco que intentó robar una casa.

Los errores de estos algoritmos son notables: el hombre de 50 años volvió a cometer un robo mayor luego de ser liberado, mientras que el chico negro que aparentemente era el doble de peligroso, no volvió a cometer un delito; el joven que resistió el arresto nunca tuvo otro problema con la policía, mientras que el blanco luego fue arrestado otras tres veces por posesión de drogas. En defensa de COMPAS podría alegarse que estamos contrastando casos extremos, pero un análisis detallado de 7000 casos mostró que el algoritmo tiende a dar factores de riesgo más altos a los negros que a los blancos: de aquellas personas que no volvieron a reincidir, el doble de negros fue considerado de alto riesgo comparado con los blancos y, por otro lado, de aquellas que sí reincidieron, casi el doble de los blancos fue considerado de bajo riesgo en comparación con los negros[114]. En principio, podemos pensar que estas consideraciones quizás no sean tan importantes porque en el fondo las decisiones las toman los jueces. Puede que sea cierto en algunos casos, pero en otros claramente no.

Paul Zilly fue arrestado por robar una cortadora de pasto y demás herramientas. La fiscalía y la defensa acordaron una pena de un año en prisión, pero al ver el juez los reportes del factor de riesgo de Zilly, inmediatamente revocó el acuerdo y duplicó la condena. Tiempo después, un defensor oficial apeló la sentencia y llamó de testigo al creador del algoritmo, quien declaró que la idea original de COMPAS

era reducir sentencias y la sobrepoblación de las cárceles, y no de imponer o siquiera sugerir castigos. Tras escuchar este testimonio, el juez redujo la condena y en su alegato admitió que, de no haber visto el reporte de COMPAS, le habría dado a Zilly una condena de medio año, o un año a lo sumo.

Un caso todavía más renombrado es el de Eric Loomis, arrestado por huir de la policía en un auto robado. Loomis fue sentenciado a ocho años y medio de prisión, pero lo más notable es que el juez explícitamente alegó en su fallo que en parte basó su decisión en el alto factor de riesgo dado por COMPAS. La gran diferencia es que en este caso el juez hizo explícito lo que era un secreto a voces. Los abogados de Loomis consecuentemente argumentaron que era injusto juzgarlo usando información provista por un algoritmo al cual no tenían acceso. Después de todo, no se sabe cómo el algoritmo de COMPAS adjudica el factor de riesgo, y todo condenado tiene derecho a acceder a la información en la que se basa su condena. La situación se torna aún más bizarra porque la compañía que provee COMPAS argumentó que el algoritmo es secreto y está resguardado por las leyes de propiedad intelectual. O sea, un algoritmo influye en la condena de una persona y esta no tiene derecho a saber cómo se efectúan estos cálculos porque se prioriza el beneficio que tiene la compañía de lucrar con la venta del algoritmo. Parece una ironía, pero la Corte Suprema del estado de Wisconsin rechazó la apelación de Loomis y la sentencia luego quedaría firme, al negarse la Corte Suprema de Estados Unidos a revisar el caso.

Viajando en Estados Unidos, me pasó que en dos o tres vuelos seguidos me revisaron todo el equipaje. No es que me molestara demasiado ya que no tenía nada que esconder, pero como en esos tiempos solía llegar muy apurado al aeropuerto, la demora inesperada agregaba bastante estrés a mis viajes. En algún momento me di cuenta de que en estos vuelos mi tarjeta de embarque tenía impreso en la parte superior el código “SSSS”. Que tonto, pensé: si me dan una tarjeta de embarque con una cuádruple S ya sé que me van a revisar y si quisiera pasar algo ilegal, desistiría de hacerlo. Pero por otro lado también temí algo mucho más alarmante: por algún motivo que me

era desconocido, quizás había entrado en alguna lista negra. Es cierto que a veces eligen al azar pasajeros a los que les hacen chequeos más exhaustivos, pero el hecho de que me había tocado dos o tres veces seguidas ya me tenía preocupado. Por suerte no volvió a pasar. Fue efectivamente coincidencia (o el hecho de que no le haya gustado mi cara a la persona con la que hice el check-in), pero me dejo pensando: sería muy injusto terminar en una lista negra de este tipo sin saber por qué.

Nos hemos referido a los algoritmos que evalúan la probabilidad de reincidencia de personas que cometieron un delito y de cómo tales evaluaciones están muy sesgadas según sea la condición social y hasta el color de piel del sujeto en cuestión. Es injusto, pero al menos nos queda el tonto consuelo (ciertamente muy tonto y egoísta) de que a nosotros no nos va a pasar si nunca hicimos nada malo. Bueno, en realidad, ni siquiera eso.

Lo aberrante de *Minority Report* es que se juzga a personas inocentes que todavía no han cometido un delito. Sorprendentemente, algo parecido está pasando en nuestros días y no hablo de que alguien me haya puesto SSSS en mi tarjeta de embarque porque ando a medio afeitar. Me refiero a algoritmos que ya se usan en ciudades como Chicago y Kansas para identificar sujetos que probablemente puedan estar involucrados en delitos[115]. Estamos hablando de gente que es juzgada a partir de su condición social, del lugar en donde vive, su entorno familiar y hasta sus contactos en las redes sociales. El argumento que se esgrime para justificar el uso de estos algoritmos es que ayudan a prevenir el crimen y, así como pueden identificarse zonas de alto riesgo para reforzar la presencia policial, también pueden identificarse personas de alto riesgo para tratar de disuadirlas —como efectivamente lo hacen— de cometer un delito. Pero pongámonos en el lugar de esta gente. Por un lado, parece útil que tengan charlas y lecciones sobre las consecuencias de delinquir, pero por el otro es injusto discriminarlos, ponerlos bajo la lupa, obligarlos a atender sesiones de asesoramiento o tener a la policía vigilándolos de cerca aunque no hayan hecho absolutamente nada. Si bien es cierto que no se los condena de antemano como en *Minority Report*, la típica

situación social marginal en la que viven se ve agravada por la discriminación de aparecer en una lista negra que, encima de todo, es generada por un algoritmo que nadie sabe cómo funciona.

Antes argumentamos que la ciencia no da seguridad sobre la posibilidad de que alguien cometa un delito. Esta es la gran diferencia con *Minority Report*: en la vida real, nunca estamos seguros. ¿Pero y si lo estuviéramos? ¿Si supiéramos con total certeza que un asesino volverá a matar ni bien quede en libertad, que haríamos? Pongámonos por un momento en la posición del juez; imaginemos que somos nosotros quienes debemos tomar la decisión. El asesino ya cumplió su condena y sería injusto retenerlo más allá de lo establecido. Por más que ha sido culpable antes, lo estaríamos castigando, como en *Minority Report*, por un delito que aún no ha cometido. Pero tampoco queremos liberarlo sabiendo que va a matar. Claramente hay un conflicto de derechos e intereses: el de proteger a las potenciales víctimas, a la sociedad en general, y el de reinserción del condenado. Ahora bien, no existe un 100% de certeza en la realidad, ¿pero si la certeza fuera de un 99%, un 95%, o quizás solo un 90%? ¿Dónde pondríamos el límite?

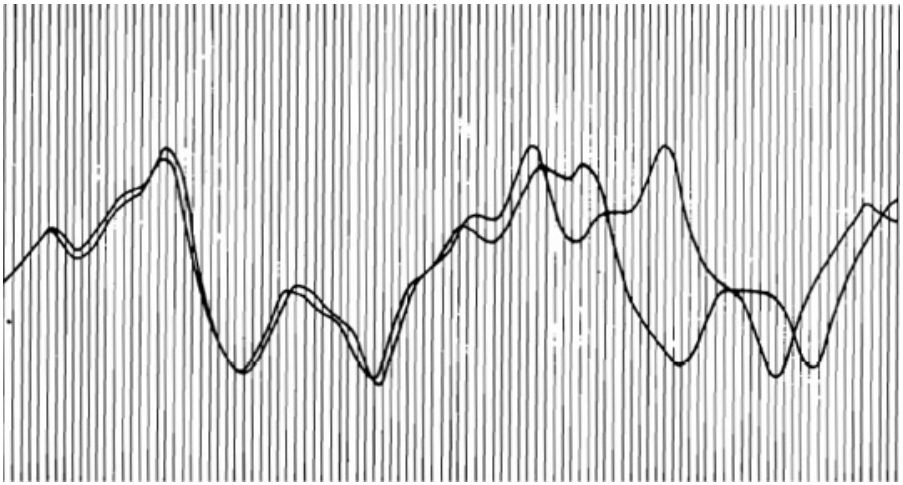
Este no es solo un problema filosófico sino algo muy cotidiano, porque la posibilidad de reincidencia de, por ejemplo, un violador o un pedófilo es altísima. La solución que la legislación de algunos países implementa es la de tener institutos especializados que sean de reclusión y rehabilitación, pero no de castigo. O sea, no es posible liberar a un criminal que constituya un peligro mayor para la sociedad, pero tampoco se lo puede seguir castigando más allá de su condena. La situación quizás no sea del todo justa para el condenado, ya que lo continuamos privando de su libertad. Pero también debemos considerar que al menos tuvo la opción de no cometer un delito en el primer caso, mientras que la eventual víctima no elige ser víctima y primero deberían atenderse sus derechos.

Estos temas son sin duda muy controversiales, pero queda pendiente la discusión más profunda y probablemente más desoladora. Aún debemos plantearnos si realmente somos libres, porque si la neurociencia nos lleva a concluir que no lo somos, en principio nada

parecería tener sentido. La solución viene dada por un descubrimiento fortuito que cambiaría la historia de la Física.

A principios de la década del 60, Edward Lorenz, un matemático y meteorólogo estadounidense, jugaba a ser el demonio de Laplace. Para ello usaba una computadora que ocupaba gran parte de su oficina, en la cual corría modelos matemáticos para tratar de hacer predicciones meteorológicas. Así como ya era posible determinar con bastante certeza la posición futura de cuerpos celestes, el desafío de Lorenz era desarrollar modelos sofisticados que pudieran predecir el clima con la mayor anticipación posible; una tarea por cierto nada fácil ya que, a pesar de también ser un sistema determinista, la dinámica de la atmósfera es muchísimo más compleja que la del movimiento de los planetas.

La computadora, una Royal McBee LGP-30, imprimía cada tanto una serie de números reportando el progreso de las simulaciones, los cuales Lorenz interpretaba para hacer sus predicciones climatológicas. Un buen día decidió continuar una simulación previa. Pero en lugar de empezar todo de nuevo, para ahorrar tiempo arrancó por la mitad, introduciendo los resultados obtenidos en la simulación anterior. Por ejemplo, si antes había hecho una predicción de 10 días, esta vez repitió la misma simulación pero arrancando a partir de los resultados del día 5. En principio, los resultados de los días 5 al 10 deberían ser los mismos ya que era exactamente el mismo modelo. Pero eso no fue lo que pasó. Los resultados coincidieron al principio, pero a medida que pasaba el tiempo eran cada vez más diferentes.



Computadora Royal McBee LGP-30 como la usada por Lorenz (arriba) y series de predicciones de Lorenz, que coinciden al principio, pero van divergiendo a medida que transcurre el tiempo (imagen de los archivos de Lorenz de 1961).

Asombrado por tal resultado, Lorenz comenzó a investigar el porqué de esta divergencia. Podía ser una falla en la computadora, lo cual no era poco habitual, o algún error al introducir los datos. Sin embargo, la computadora funcionaba perfectamente y los datos eran correctos. La diferencia entre ambas simulaciones se debía a algo mucho más sutil; un accidente que cambiaría la historia de la ciencia y que daría lugar a lo que hoy se conoce como *teoría del caos*[116].

El problema era que la computadora realizaba los cálculos con una precisión de 6 dígitos decimales, pero solo imprimía 3. Por ejemplo, la computadora operaba con el número 0,506127, pero daba el resultado 0,506. Al repetir la simulación, Lorenz introdujo los resultados con una precisión de 3 decimales, pero el error mínimo de no considerar los decimales a partir de la tercera cifra después de la coma, hizo divergir rápidamente ambas simulaciones. Y esta es justamente la característica principal de los sistemas complejos o caóticos[117]: diferencias mínimas en las condiciones iniciales se agrandan exponencialmente. Por este motivo los sistemas caóticos son impredecibles.

Supongamos que tenemos que introducir datos en un modelo para hacer estimaciones climatológicas (por ejemplo, la presión,

temperatura y humedad medida en distintos lugares). ¿Cuántos dígitos decimales debemos considerar? Lo que dice la teoría del caos es que por más grande que sea la precisión que usemos con un sistema caótico, por más mínimo que sea el error, en algún momento el sistema se va a volver totalmente imprevisible[118]. Es por eso que en nuestros días, más de medio siglo después de los modelos de Lorenz, seguimos sin poder predecir el tiempo más allá de unos pocos días.

La sensibilidad a las condiciones iniciales es lo que popularmente se conoce como el *efecto de la mariposa*; según la famosa frase de Lorenz, el aleteo de una mariposa en Brasil influye en un tornado en Texas. Por supuesto que esta influencia es infinitesimal y otra infinidad de factores tendrán un impacto mucho más directo en el clima de Texas. Pero la idea es que ningún modelo de un sistema caótico, por más refinado que sea, podrá considerar efectos como el aleteo de la mariposa en Brasil y tarde o temprano esto hará que las predicciones sean inútiles. La teoría del caos cambia radicalmente la visión del determinismo de Laplace y de la mecánica clásica. El gran postulado tras la revolución científica de que, en principio, podría predecirse la evolución del universo ya no es cierto; hay una gran limitación infranqueable.

Por lo tanto, aun siendo válido el determinismo, los sistemas complejos y caóticos son impredecibles. Evidentemente uno de estos sistemas es el cerebro. Miles de millones de neuronas disparan de una u otra manera generando nuestras acciones, nuestros pensamientos, cada una de las decisiones que tomamos a lo largo de nuestra vida. Y si bien es cierto que nuestras decisiones están inexorablemente determinadas por la actividad de las neuronas y que dicha actividad está a su vez determinada por estados anteriores (e infinidad de otros factores), la clave es que nuestro comportamiento es impredecible. Vimos que haciendo uso de metodologías avanzadas podemos predecir y hasta influenciar si una persona va a mover un dedo, pero por más avanzadas que sean estas tecnologías, nunca sabremos con certeza si va a proponerle casamiento a su novia, estudiar arquitectura o ir al gimnasio esa misma tarde.

La teoría del caos es lo que nos rescata de la gran desolación que

genera la idea del determinismo. Si bien nuestros actos están determinados y siguen reglas físicas, son impredecibles —no existe ni existirá nadie que, como en *Minority Report*, pueda vislumbrar nuestros actos futuros y juzgarnos por ellos. Pero probablemente aún nos quede un sinsabor, algo que no nos termina de cerrar del todo: el hecho de que pareciera que no tenemos control, pues en el fondo es el disparo de nuestras neuronas (y no nosotros) lo que determina nuestras decisiones. En otras palabras, no poseemos, como postulara Descartes, una mente autónoma que nos lleva a hacer uso de nuestra voluntad para elegir una u otra cosa. Pero ahondemos un poco en esta idea. ¿Cómo haría una mente autónoma para tomar decisiones? Si no hiciera más que pesar evidencia —es decir, poniendo en la balanza todas las razones a favor y en contra— por más que seamos inconscientes de muchas de esas razones, la decisión sería igualmente un proceso determinista.

La noción ingenua que tenemos del libre albedrío es, en cambio, que efectivamente consideramos evidencia pero que hay algo más que nos lleva a ejercer nuestra voluntad. Pero ¿qué es ese algo? ¿Quizás rasgos de nuestra personalidad, como el hecho de que somos honestos, emocionales o aguerridos? Sin embargo, estos factores pueden explicarse por nuestra herencia genética y nuestra formación, también siguiendo reglas deterministas.

Lo opuesto al determinismo es la aleatoriedad; no hay más opciones. Si el determinismo no explica nuestras decisiones, entonces las tomamos al azar, tirando una moneda, o tal vez haciendo una mezcla de las dos cosas. Pero aun admitiendo, como postulara Aristóteles, una cuota de azar para librarnos del determinismo —o sea, teniendo una predisposición a elegir algo pero sabiendo que por chance podemos terminar eligiendo algo distinto— la idea de que, al menos en parte, elegimos al azar, también dista mucho de la noción de libre albedrío de Descartes.

En conclusión, somos libres, ejercemos nuestra voluntad. Mas el libre albedrío no es algo mágico e inexplicable, sino que está dado por el comportamiento de nuestras neuronas. La falla radica en pensar que son nuestras neuronas y no nosotros los que estamos a cargo de

nuestras decisiones, pues al hacer esta afirmación caemos de nuevo en la trampa del dualismo cartesiano, tan arraigado en nuestra intuición y tan difícil de evitar. Nosotros y nuestras neuronas son una misma cosa. Si mis neuronas disparan de una manera específica para decidir sobre lo que escribo en este momento, soy yo quien lo está haciendo. No existe la mente autónoma que toma decisiones libremente, ejerciendo su voluntad, pero sí existen neuronas y procesos conscientes e inconscientes —nuestro Yo— que nos llevan a hacer una u otra cosa. Por más que la idea ingenua que tenemos sobre el libre albedrío, aquella de Descartes, sea una ilusión, aún somos libres.

En el fondo, nos rasgamos las vestiduras por algo que, pensándolo un poco, parecería obvio. Realmente no hace falta conocer los experimentos de Libet para concluir que la noción del libre albedrío de Descartes no tiene sentido. Tampoco hace falta saber mucho de neurociencia para percatarnos de lo frágiles y maleables que son nuestras decisiones. Solo nos basta un poco de introspección para ver que en muchos casos tendemos a elegir impulsivamente y que a posteriori racionalizamos estas decisiones. Optamos por un trabajo y no por otro porque, quizás incluso inconscientemente, nos cayó bien el apretón de manos de nuestro futuro jefe o apreciamos el sillón reclinable de nuestra nueva oficina. Pero indudablemente después racionalizamos y argumentamos que el trabajo elegido es en realidad el más apropiado por uno u otro motivo.

Un ejemplo contundente de cuán frágiles y maleables son nuestras decisiones está dado por lo que se conoce como *choice blindness* (ceguera de elección). En uno de estos experimentos se mostró a distintas personas pares de fotos de mujeres desconocidas. Tras ver cada par, los sujetos indicaron cuál de las dos mujeres les parecía más atractiva; luego el experimentador les mostraba la foto elegida preguntándoles por qué se habían decidido por ella. Lo interesante es que, usando una adaptación de un truco de magia, el experimentador cada tanto intercambiaba la foto señalada sin que los sujetos se dieran cuenta y estos terminaban explicando por qué les gustaba justamente la persona que no habían elegido[119].

Siguiendo con esta idea, el mismo grupo de investigadores mostró

que incluso podían cambiar tendencias políticas de la gente: durante el período de una elección general en Suecia, manipularon (también usando una suerte de truco de magia) las respuestas escritas que personas escogidas al azar daban a un cuestionario sobre actitudes políticas y observaron que minutos después de haber hecho el cuestionario, los sujetos justificaban actitudes políticas muy distintas a las que inicialmente habían manifestado en sus respuestas[120]. Estos experimentos no hacen más que reafirmar el hecho de que tendemos a elegir rápida e impulsivamente y que luego construimos un andamiaje de argumentos para justificar esas decisiones, sosteniendo así la idea de que elegimos basándonos en el uso de nuestro razonamiento y voluntad; la ilusión del libre albedrío.

97 El determinismo fue luego dejado de lado a partir del surgimiento de la mecánica cuántica, a principios del siglo XX. Según el principio de incerteza de Heissenberg —uno de los pilares de la mecánica cuántica— no podemos conocer al mismo tiempo la posición y la velocidad exacta de una partícula. Entonces, la noción del genio de Laplace no es posible, y no por una cuestión práctica, sino porque violaría las leyes de la física, las cuales limitan nuestro conocimiento del estado de las partículas que componen el universo. Pero más allá de esta crítica al determinismo, los argumentos en contra del libre albedrío siguen siendo válidos, ya que la incerteza en la mecánica cuántica se da a escalas microscópicas, demasiado pequeñas como para influir en el comportamiento de las neuronas.

La idea de que efectos cuánticos pueden explicar comportamientos del cerebro fue, sin embargo, propuesta por el aclamado matemático Roger Penrose en su libro *The emperor's new mind* (Oxford University Press, 1989). Aunque esta idea, proclamada sin ningún tipo de fundamento científico —casi como diciendo: si no entendemos cómo funciona el cerebro, debe haber algún efecto cuántico— fue, salvo raras excepciones, unánimemente refutada por la comunidad científica. Para más detalles, ver: Christof Koch and Klaus Hepp. "Quantum mechanics in the brain". *Nature* 440: 611-612; 2006.

98 Libet B, Gleason C, Wright E and Pearl D. "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act". *Brain* 106:623-642; 1983.

99 Para una discusión crítica del experimento de Libet ver: Benjamin Libet. "Unconscious cerebral initiative and the role of conscious will in voluntary action".

Behavioral and Brain Sciences 8: 528-566; 1985.

100 Itzhak Fried, Amiram Katz, Gregory McCarty, Kimberlee Sass, Peter Williamson, Susan Spencer and Denis Spencer. "Functional organization of human supplementary motor cortex studied by electrical stimulation". *The Journal of Neuroscience* 11: 3656-3666; 1991.

101 Figura adaptada de: Wilder Penfield. "Some mechanisms of consciousness discovered during electrical stimulation of the brain". *Proceedings of the National Academy of Sciences USA* 44: 51-66; 1958.

102 Joaquim Brasil-Neto, Álvaro Pascual-Leone, Josep Valls-Sole, Leonardo Cohen and Mark Hallett. "Focal transcranial magnetic stimulation and response bias in a forced-choice task". *Journal of Neurology, Neurosurgery and Psychiatry* 55: 964-966; 1992.

103 Sergio Della Sala, un neurólogo especialista en el síndrome de la mano ajena, suele referir a Cortázar y a Peter Sellers en sus publicaciones. Ver, por ejemplo: Sergio Della Sala. "The anarchic hand". *The Psychologist* 18: 606-609; 2005.

104 Para una descripción de estos casos ver, por ejemplo:

- D. Geschwind, M. Iacoboni, M. Mega, D. Zaidel, T. Cloughesy and E. Zaidel. "Alien hand syndrome: Interhemispheric motor disconnection due to a lesion in the midbody of the corpus callosum". *Neurology* 45: 802-808; 1995.
- G. Goldberg, N. Mayer and J. Toglia. "Medial frontal cortex infarction and the alien hand sign". *Archives of Neurology* 38: 683-686; 1981.

105 La conciencia que precede a nuestros actos no solo nos da la sensación de estar en control, sino que también nos ofrece un mecanismo de veto para poder desistir de hacer una acción antes de ejecutarla. Ver, por ejemplo: Marcel Brass and Patrick Hagard. "To do or not to do: The neural signature of self-control". *The Journal of Neuroscience* 27:9141-9145; 2007.

106 Sarah-Jayne Blakemore, Daniel Wolpert and Christopher Frith. "Abnormalities in the awareness of action". *Trends in Cognitive Science* 6: 237-242; 2002.

107 P. McGuire, G. Shah and R. Murray. "Increased blood flow in Broca's area during auditory hallucinations in schizophrenia". *Lancet* 342: 705-706; 1993.

108 Sean Spence. "Free-will in the light of neuropsychiatry". *Philosophy, psychiatry and psychology* 3: 75-90; 1996.

109 Sarah-Jayne Blakemore, Christopher Frith and Daniel Wolpert. "Spatio-temporal prediction modulates the perception of self-produced stimuli". *Journal of Cognitive Neuroscience* 11: 551-559; 1999.

110 Una representación brillante y emotiva de lo que implica el síndrome de Tourette está dada en el corto *Front of the class* (puede encontrarse fácilmente en YouTube), el cual relata los problemas sufridos por Brad Cohen, un chico con Tourette que era constantemente discriminado en la escuela y que actualmente se desempeña como maestro, deseando ser la clase de maestro que él mismo hubiera querido tener.

111 Estos resultados están descriptos en: Kathleen Vohs and Jonathan Schooler.

“The value of believing in free will: encouraging a belief in determinism increases cheating”. *Psychological Science* 19: 49-54; 2008.

112 Su búsqueda para dilucidar si efectivamente era un psicópata es el tema de su libro *The Psychopath Inside* (Penguin, 2013).

113 Michael Gazzaniga; *Who’s in charge?*; Robinson; Londres; 2012.

114 Para más detalles sobre estos casos y estadísticas ver el informe (“Machine bias”) del portal de noticias *ProPublica*.

115 El uso de estas metodologías ha sido informado en varios artículos del periódico *The New York Times*.

116 Sin entrar en detalles, pero para no ser injusto en las atribuciones, cabe aclarar que los orígenes de la teoría del caos se remontan al matemático francés Henri Poincaré, a principio del siglo XX. Para una introducción rigurosa y amena a la teoría del caos, ver el libro *Chaos: Making a New Science* (Viking, 1987) de James Gleick.

117 Aclaro que matemáticamente no es lo mismo un sistema complejo que uno caótico, pero en el contexto de esta discusión no vale la pena ahondar en esas diferencias técnicas y usaré ambos términos con el mismo sentido.

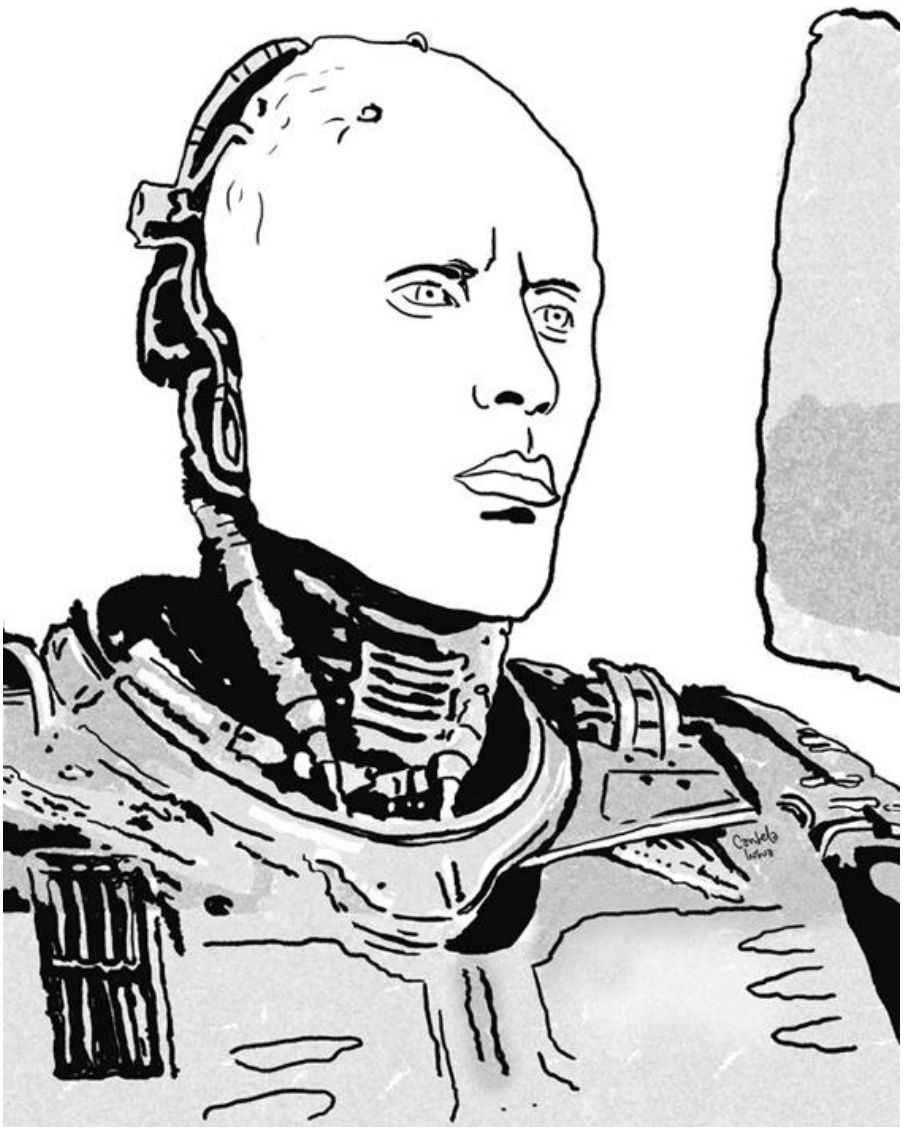
118 El *exponente de Lyapunov* determina el grado de sensibilidad a las condiciones iniciales o, en otras palabras, qué tan rápido un sistema caótico se vuelve imprevisible.

119 Peter Johansson, Lars Hall, Sverker Sikstrom and Andreas Olsson. “Failure to detect mismatches between intention and outcome in a simple decision task”. *Science* 310:116-119; 2005.

120 Este estudio esta descripto en: Lars Hall, Thomas Strandberg, Philip Pärnamets, Andreas Lind, Betty Tärning and Petter Johansson. “How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions”. *PLOS* One 8(4): e60554; 2013.

VII. Robocop

Cyborgs e identidad



Cuando era chico, esperaba ansioso cada episodio de *The six million dollar man*, la serie de televisión que en Argentina se conoció como *El hombre nuclear*, en donde Steve Austin, astronauta y piloto de prueba, sufre un grave accidente al estrellarse la nave en la que volaba. Una voz en *off* que aparece al comienzo resume la trama de la serie: "Steve Austin, astronauta. Su vida está en peligro. Lo reconstruiremos. Poseemos la tecnología para convertirlo en un organismo cibernético, poderoso, superdotado"[121]. La reconstrucción cuesta 6 millones de dólares y consiste en reemplazarle un ojo, con el que puede ver 20 veces más lejos que el ojo humano, ambas piernas, que le permiten correr a velocidades superiores a los 100 kilómetros por hora, y el brazo derecho, con el cual posee una fuerza extraordinaria. No recuerdo el argumento de ninguno de los capítulos, pero sí recuerdo vagamente escenas en donde Austin corría a la velocidad de un auto, saltaba muros o miraba lo que ocurría a la distancia.

En la década del 70, la serie fue un éxito. Estaba basada en la novela *Cyborg*, de Michael Caidin, y esto era Austin, un cyborg (abreviatura de *cybernetic organism*: organismo cibernético), un humano con partes biomecánicas; un híbrido entre hombre y máquina. Como muchos de mi edad, crecí viendo *El hombre nuclear* como algo fascinante e inalcanzable. Pero hoy estamos rodeados de cyborgs —personas con implantes de cóclea, órganos artificiales o diminutos marcapasos— y no parece tan lejano poder reemplazar el brazo, las piernas o el ojo de una persona por prótesis sintéticas. Una idea análoga da pie a la película *Robocop*, en la que un policía, Alex Murphy, es reconstruido después de sufrir lesiones graves y continúa cumpliendo su función, pero haciendo uso de las facultades extraordinarias que le proporcionan sus partes biomecánicas.

En el capítulo anterior discutimos la posibilidad de anticiparse a los delitos y eventualmente prevenirlos. En el caso de *Robocop*, la idea es

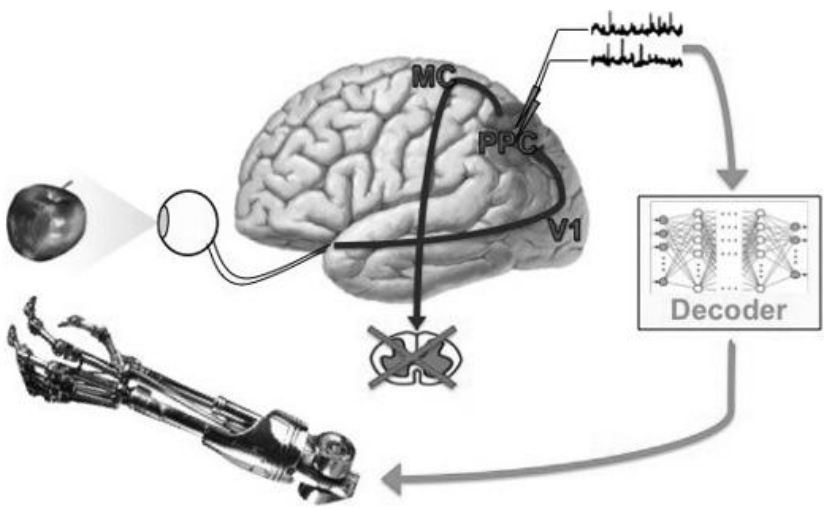
bajar la tasa de delitos a partir de tener policías casi imbatibles para enfrentar a los criminales. Un aspecto interesante de la película es que plantea la posibilidad de usar robots (en vez de cyborgs) para esta función. Pero más allá de los avances espectaculares de la Inteligencia Artificial (ver capítulo 1), en principio es preferible que haya un humano en control, pues, aunque sea un cyborg, es su mente humana la que toma decisiones usando el sentido común y las emociones, algo que las máquinas (todavía) no poseen. Las posibles consecuencias de una falla en la programación de un robot portando armas y haciendo uso de la fuerza son aterradoras. El comienzo de la película muestra justamente cómo un robot de seguridad malinterpreta las intenciones de una persona inocente y la acribilla a balazos (lo cual lleva a decidir el uso de cyborgs).

El hecho de tener un humano en control impone desafíos importantes. Ya no es tan difícil construir un brazo artificial que replique la funcionalidad del brazo humano. Lo realmente complejo es manejar el brazo mecánico directamente a partir de la actividad del cerebro, es decir, leer las intenciones de movimientos del cerebro y transmitirlas a un brazo o a cualquier otro miembro biomecánico. Este tipo de desarrollos es lo que se conoce como *neuroprostética*, un área de investigación de muchísimo auge en los últimos años[122].

El objetivo primario de los avances en este campo no es crear cyborgs como Robocop o Steve Austin, sino desarrollar herramientas con las que pacientes tetrapléjicos (o con amputaciones) puedan comandar miembros prostéticos para alcanzar objetos, alimentarse por sí mismos o volver a caminar. Esa es la principal línea de investigación del laboratorio de Richard Andersen quien, junto a Christof Koch, fue mi mentor en Caltech. Recuerdo quedar atónito al llegar por primera vez a su laboratorio y ver de primera mano que la posibilidad de usar implantes neuronales acoplados a miembros prostéticos no estaba lejos de ser realidad y que la idea de *El hombre nuclear* ya no era una quimera.

El mayor logro científico de Richard es haber descubierto dos áreas en la corteza parietal posterior, cuya función es la de planear la intención de distintos movimientos (de los brazos y de los ojos)[123]. La

corteza parietal posterior tiene una ubicación privilegiada para esta función, pues se encuentra entre la corteza visual primaria —en donde se codifica la visión de, por ejemplo, una manzana— y la corteza motora, donde se planea la ejecución del movimiento para alcanzarla (ver figura). A partir de esto, se desprende naturalmente la idea de las prótesis neuronales. Un paciente como el actor Christopher Reeve —quien luego de personificar a *Superman* irónicamente terminó postrado en una silla de ruedas, tras una lesión en la espina dorsal al caer de su caballo[124]— puede ver el objeto que quiere alcanzar y puede también planear el movimiento para alcanzarlo. Es decir, la intención del movimiento (codificada en la corteza parietal posterior y en la corteza motora) es análoga a la de una persona normal; el problema es que el daño en la espina dorsal no le permite transmitir esta señal a los músculos del brazo para ejecutarlo. Entonces, la prótesis neuronal lee la actividad de neuronas con electrodos implantados en el cerebro, y usando algoritmos de decodificación (que ya describiéramos en capítulos anteriores) predice el movimiento que desea realizar el sujeto, para llevarlo a cabo con un miembro mecánico.



Esquema de una prótesis neuronal. Para agarrar, por ejemplo, una manzana, la visión del objeto genera activaciones en la corteza visual primaria (V1), que se transmiten a la corteza parietal posterior (PPC), en donde se genera la intención del movimiento. Estas activaciones luego se transmiten a la corteza motora (MC), en donde se implementa la ejecución del movimiento, y a los

músculos del brazo a través de la espina dorsal. En pacientes con una lesión en la espina dorsal, la señal no llega a los músculos. Sin embargo, puede decodificarse la intención del movimiento del sujeto a partir del registro de neuronas en PPC (o MC), la cual se transmite a un brazo mecánico que efectúa el movimiento.

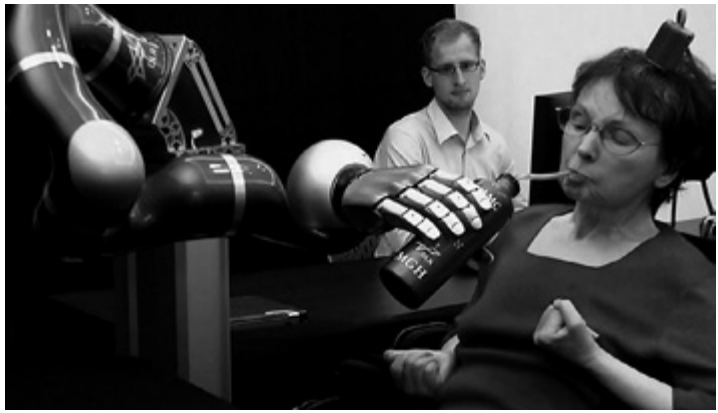
La idea del algoritmo de decodificación es similar a la descripta en el capítulo 5, donde vimos que analizando el registro de neuronas en pacientes que miran distintas fotos podíamos predecir qué foto estaban observando. En este caso, las neuronas en la corteza parietal posterior no responden a Halle Berry o Jennifer Aniston —como aquellas neuronas que encontramos en humanos— sino a la intención de realizar movimientos específicos. Por ejemplo, una de estas neuronas responderá cada vez que el sujeto planea mover el brazo hacia abajo, otra cuando planea moverlo a la izquierda, etcétera. Tomando la actividad de un par de decenas de neuronas, es entonces posible predecir a la perfección los movimientos del brazo, como fue demostrado a partir de experimentos con monos[125].

Una vez que se puede leer con precisión la intención del movimiento, solo falta activar el brazo mecánico con estas señales y hacer lo que en la jerga se conoce como *close the loop* (cerrar el circuito): que el sujeto vea el movimiento del brazo mecánico y que pueda controlarlo mediante la actividad de sus neuronas, corrigiendo eventuales errores a partir del *feedback* visual. La gran ventaja es que no hace falta decodificar los detalles específicos del movimiento; de ello se encarga la robótica. Si el cerebro da la instrucción de mover el brazo a un punto determinado, son los algoritmos de control del brazo mecánico (y no el cerebro) quienes se ocupan de implementar un movimiento fluido hacia este punto, aplicando coordinadamente distintas fuerzas a cada una de las articulaciones. Esto simplifica muchísimo el problema, pues solo tenemos que decodificar instrucciones generales y no los detalles de cómo implementarlas (así como solo somos conscientes de querer mover el brazo a determinado lugar y no de las distintas fuerzas que debemos aplicar a cada uno de los músculos involucrados para realizar el movimiento).

Uno de los resultados más espectaculares en este campo lo obtuvo el

grupo de Andy Schwartz, en la Universidad de Pittsburgh, mostrando que un mono era capaz de controlar continuamente el movimiento de un brazo mecánico a partir de la actividad de un par de decenas de neuronas. Tal era la precisión con la que el mono controlaba el brazo mecánico, que podía agarrar y llevar a la boca pequeños trozos de comida colocados en distintos lugares enfrente de él[126]. Este resultado no solo causó un gran revuelo entre los científicos, sino que también fue ampliamente difundido por la prensa[127]. La conclusión era obvia, si Schwartz y sus colegas habían logrado un avance tan notable en monos, ¿podrían implementarse tecnologías análogas en humanos?

La respuesta no se hizo esperar mucho. Pocos años después, la comunidad científica (y el público en general, a partir de una impresionante divulgación en los medios), era testigo de cómo resultados parecidos eran obtenidos con pacientes tetrapléjicos[128]. La imagen muestra justamente a uno de estos pacientes, quien logró manipular un recipiente para beber su contenido, marcando un hito en la historia de la neuroprostética.



Mono tomando un malvavisco con un brazo mecánico controlado por su cerebro, que luego llevará a su boca (izquierda). Paciente tetrapléjico manejando un brazo mecánico a partir de la actividad de sus neuronas para tomar una bebida (derecha). Adaptado de Velliste et al., *Nature*; 2008 y Hochberg et al., *Nature*; 2012.

Los avances descriptos aun distan mucho de Steve Austin o Robocop, pero constituyen un *proof of principle*, una demostración fehaciente de la posibilidad de lograr que una persona controle prótesis biomecánicas directamente con su cerebro. Más ¿qué tan lejos estamos de desarrollar cyborgs sofisticados como los de las películas? ¿Siguen tan solo siendo parte del imaginario colectivo, o tendremos a estos cyborgs caminando por la calle en un futuro cercano?

Un aspecto importante que ha demorado un progreso más marcado es meramente ético y técnico, ya no tanto científico. Este tipo de interfaces neuronales requiere de una cirugía mayor para implantar decenas o cientos de electrodos en el cerebro. La gran pregunta es entonces: ¿se justifica esta cirugía, con los eventuales riesgos que implica?

Ya vimos que los avances de la robótica permiten realizar tareas relativamente complejas (como alcanzar una bebida y llevarla a la boca) y si pudiéramos garantizarles a los pacientes que a partir de la implantación de electrodos podrían hacer ese tipo de tareas por el resto de sus días, entonces la intervención quirúrgica estaría ampliamente justificada[129]. Por el contrario, si la cirugía solo permitiera comandar las prótesis neuronales por un tiempo

relativamente limitado, parecería no tener sentido. (Aunque esta discusión también tiene sus matices, pues aún en el caso de no hacer diferencia en el paciente implantado, podría también justificarse la intervención quirúrgica si diera lugar a avances que eventualmente permitieran mejorar la calidad de vida de otros pacientes en el futuro.)

El problema es que, si bien es cierto que a partir de registrar la actividad de decenas o cientos de neuronas es factible manipular una prótesis neuronal, todavía no es posible mantener registros estables de estas neuronas que duren varios años. En los primeros pacientes tetrapléjicos implantados con electrodos, solo pudieron mantenerse registros neuronales por un período acotado. Esto fue obviamente decepcionante para los pacientes, ya que al decaer la calidad de los registros perdieron la capacidad de comandar el brazo mecánico que antes controlaban.

Resultados actuales son mucho más promisorios, llegando incluso a mantener los registros por varios años[130]. Sin embargo, estos registros no son estables. O sea, ya pueden registrarse decenas de neuronas durante años, pero las neuronas que se registran van cambiando; día a día, algunas se pierden y otras nuevas aparecen[131]. ¿Cómo es entonces posible que funcione un algoritmo de decodificación? Imaginemos que entrenamos el algoritmo teniendo en cuenta la actividad de una neurona codificando movimientos hacia arriba. Cada vez que dispara esta neurona, el algoritmo interpretará que el paciente quiere mover el brazo hacia arriba. Pero obviamente el algoritmo va a confundirse si tiempo después varían las condiciones de los registros, haciendo que en lugar de esta neurona aparezca otra que codifica movimientos a la izquierda.

La solución a este problema viene (al menos en parte) dada por la *plasticidad neuronal*: las neuronas adaptan su actividad para lograr que el sistema funcione. Siguiendo con el ejemplo anterior (y simplificando un poco para que la idea quede clara), si una neurona codifica movimientos hacia arriba y es reemplazada por otra que codifica movimientos hacia la izquierda, esta nueva neurona podrá modular su actividad (por ejemplo, aumentando la codificación de

movimientos hacia arriba y reduciendo la de movimientos hacia la izquierda) para lograr que el brazo se mueva coherentemente. Este tipo de adaptación neuronal fue mostrada en un experimento en el cual primero entrenaron a un mono para que pudiera mover un cursor en la pantalla (usando un algoritmo de decodificación basado en la actividad de las neuronas), luego perturbaron el algoritmo, cambiando la información dada por un grupo de las neuronas registradas, y finalmente observaron que algunas neuronas adaptaban sus repuestas para compensar la perturbación, con lo cual el mono podía seguir moviendo coherentemente el cursor en la pantalla[132].

En conclusión, la plasticidad neuronal (y el uso de algoritmos de decodificación adaptativos[133]) permiten lidiar con la variabilidad de los registros y lograr que las prótesis neuronales duren varios años[134]. Aunque si nos planteamos seriamente hacer realidad desarrollos como los de *Robocop* o *El hombre nuclear*, todavía nos quedan por discutir aspectos no menos importantes.

La imagen que ilustra este capítulo muestra el momento en que Alex Murphy (o Robocop) se quita el casco y se ve tal cual es: un cyborg. Tras examinar un tiempo su imagen reflejada, en parte humana y en parte robot, pregunta por “la familia de Murphy”. Lo interesante es que no pregunta por su familia sino por “la de Murphy”, en tercera persona, como si preguntara por la familia de otro.

Robocop fue estrenada en 1987 y en una remake de 2014, una escena análoga es mucho más dramática. Tras haber recuperado la consciencia, Murphy le pregunta al doctor Dennett Norton[135], a cargo de la operación, por el “traje metálico” que lleva puesto. Norton le contesta que no es un traje, sino que es él mismo. Luego le muestra lo que queda de su cuerpo y Murphy ve frente a un gran espejo cómo se van desensamblando uno a uno sus miembros robóticos: las piernas, los brazos, el abdomen, el tórax. Con cada miembro que comprueba que no es parte de su cuerpo crece su desesperación y siente como si su vida se hubiera acabado, como si aquella persona que era —su propio yo— hubiese dejado de existir. Sin embargo, no había otra posibilidad de mantenerlo vivo y, a modo de consuelo, el doctor Norton le dice que es su cuerpo el que desapareció, pero que él sigue

allí. Pero ¿sigue siendo Murphy el cyborg frente al espejo?

Este es el problema de la identidad, cuya discusión se remonta a la famosa *paradoja del barco de Teseo*, descripta por Plutarco[136]. Cuenta la leyenda que el poderoso rey Minos de Creta exigía a los atenienses (a quienes culpaba de la muerte de su hijo) que todos los años le pagaran un tributo de sangre y le enviaran siete jóvenes y otras tantas doncellas, quienes eran dejados en un laberinto infranqueable para ser devorados por el Minotauro. En uno de estos viajes, Teseo, hijo del rey Egeo y héroe mitológico de Atenas, se embarca junto a los jóvenes para enfrentar al Minotauro. Al llegar a Creta, Ariadna, hija del rey Minos, se enamora de Teseo y le da un ovillo de hilo para que pueda marcar su paso y eventualmente encontrar la salida del laberinto. Teseo da muerte al Minotauro, logra escapar del laberinto y, tras haber puesto fin al brutal sacrificio de los jóvenes atenienses, emprende victorioso la vuelta a Atenas en su barco[137].

El barco de Teseo pasó a constituirse en un símbolo para los atenienses, quienes, tras haberse liberado del injusto sometimiento a manos de Creta, todos los años, y durante siglos, enviaron el barco a la isla de Delos para rendirle culto a Apolo[138]. Sin embargo, con el paso del tiempo sus tablones de madera se fueron resquebrajando y debieron ser reemplazados por tablones nuevos. La monumental pregunta que se hace entonces Plutarco es si ese barco, careciendo de los tablones originales, continúa siendo el barco de Teseo.

El planteo de Plutarco está relacionado con las *paradojas de sorites* (del griego: montón): si a un montón de arena le quitamos un grano, seguirá siendo un montón de arena; pero si repetimos este procedimiento recursivamente, quitando de a un grano a la vez, siempre diremos que seguimos teniendo un montón de arena (si N granos son un montón, N-1 granos también lo son), hasta eventualmente quedarnos sin nada. Por lo tanto, alteraciones graduales enmascaran la percepción de cambio[139].

Ilustración de la nave de Teseo. Tras siglos de mantenimiento, se van reemplazando las partes de la nave deterioradas por el tiempo. Pero ¿continúa siendo el barco de Teseo?

Volviendo al barco de Teseo, la decisión de si continúa siendo el mismo barco parecería depender de cuán rápida es la transformación. Al no estar corroídos por el paso del tiempo, los tablones nuevos se verán algo distintos, pero si cada varios años solo se reemplaza una u otra madera, tenderíamos a decir que el barco restaurado es el mismo. Por el contrario, si alguien decide reemplazar gran parte del barco de un día para el otro, costaría aceptar que sigue siendo el barco de Teseo.

Este razonamiento se aplica directamente a la discusión sobre la identidad de Robocop. Si paulatinamente se le reemplazara una pierna, tiempo después la otra, luego un brazo y así sucesivamente, en ningún momento Robocop sentiría que dejó de ser Murphy. Pero si la transformación es de golpe, como en la película, enfrentamos un problema de identidad. Por lo tanto, la continuidad del cuerpo —el hecho de que los cambios en el cuerpo se den lentamente— es lo que aparentemente define la identidad personal. Después de todo, seguimos siendo la misma persona a pesar de envejecer y de que

nuestra apariencia vaya cambiando con el paso de los años. Estos argumentos son en principio muy obvios y convincentes, pero la discusión sobre qué es lo que determina la identidad es en realidad mucho más compleja.

La noción de que la identidad personal va ligada a la continuidad del cuerpo fue desafiada por John Locke. Para ilustrar su postura, propuso la famosa *parábola del príncipe y el zapatero*: un buen día la mente del príncipe, con todas las memorias de su vida pasada, es transferida al cuerpo del zapatero. ¿Quién es quién? —se preguntó Locke[140]. ¿Pasará el zapatero a ser príncipe, ya que ahora lleva su mente? Locke argumentó que la identidad va ligada a la a la memoria, pues el príncipe seguirá reconociéndose como tal aunque se vea en un cuerpo extraño. Esta visión es radicalmente opuesta a la que postuláramos a partir de la paradoja de Teseo; lo que define a la persona no es la continuidad de su cuerpo, sino su memoria.

La metamorfosis de Kafka ofrece un ejemplo aún más contundente. Al despertar un día, Gregor Samsa nota que progresivamente va convirtiéndose en un abominable insecto. Pero Kafka continua naturalmente el relato en primera persona y no tenemos problema en asumir que, a pesar de la monstruosa deformación de su cuerpo, Gregor y el insecto continúan siendo uno y el mismo. Por el contrario, nuestra percepción de identidad es muy distinta si lo que cambia es la memoria. De alguien que sufre un devastador déficit de memoria debido a un estado avanzado de alzhéimer, decimos que "ya no es el mismo". El cuerpo no ha cambiado, pero la persona se ha ido diluyendo con el paso tiempo a partir de su olvido.

En conclusión, Robocop seguirá siendo Murphy mientras mantenga su memoria[141]. (Aunque en la película tratan de borrársela para que sea más eficiente y tome decisiones automáticamente, sin la menor vacilación provocada por sus memorias y emociones pasadas.) Por otro lado, la neurociencia también nos revela que, tras un tiempo, Murphy terminará sintiendo que las partes mecánicas implantadas son parte de su cuerpo, así como un tablón que reemplaza a otro en mal estado pasa a ser parte del barco de Teseo.

Una de las demostraciones más sencillas y a la vez más

sorprendentes en neurociencia es la *ilusión de la mano de goma*[142]: un sujeto coloca su mano bajo la mesa (o tras un tabique, fuera del alcance de su vista) y en el lugar en donde naturalmente estaría apoyada su mano sobre la mesa se coloca una mano de goma; con dos pinceles se acarician al mismo tiempo los dedos de ambas manos (la real y la de goma) y luego de un tiempo, la coherencia entre lo que la persona siente en su mano real y lo que ve en la de goma le genera la ilusión de que la mano de goma es parte de su cuerpo. Tan fuerte es la ilusión, que ni siquiera hace falta que la mano de goma se parezca a la real (basta con usar un guante para lavar platos relleno con algodón) y muchas veces suele terminarse el experimento martillando repentinamente la mano de goma, lo cual hace que el sujeto salte asustado, pues a pesar de no sentir nada, en ese momento no puede evitar creer que le han martillado su propia mano.

Ilusión de la mano de goma. Luego de un tiempo de acariciar con dos pinceles los mismos dedos en la mano real (derecha) y la de goma (izquierda), el sujeto cree que la mano de goma es parte de

su cuerpo[143].

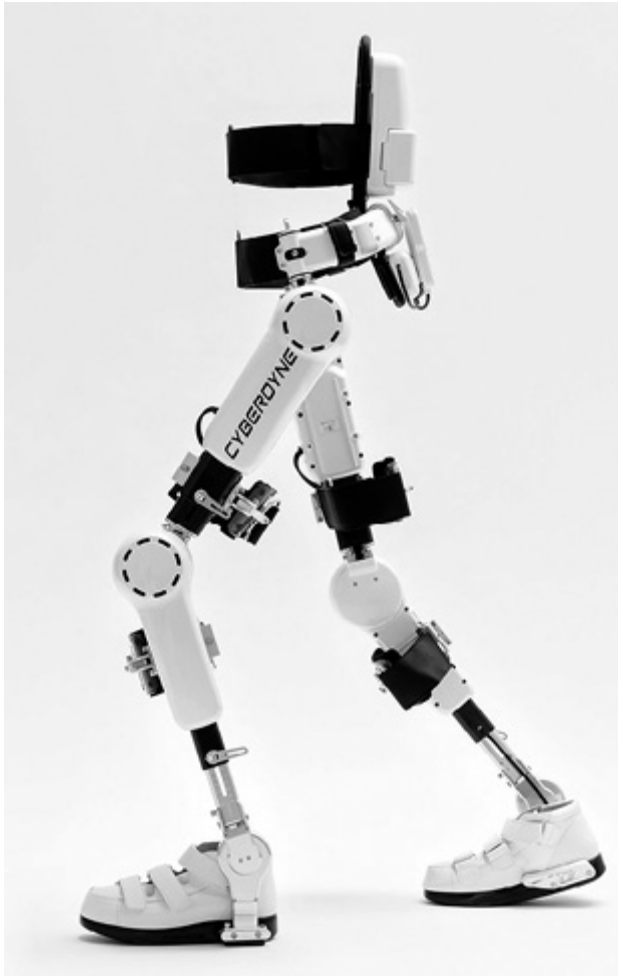
La clave para que la ilusión de la mano de goma funcione es que el estímulo táctil en la mano real sea coherente con el estímulo visual aplicado a la mano de goma. La ilusión se desvanece inmediatamente si se estimulan dedos distintos o si no se estimulan las dos manos al mismo tiempo, porque las expectativas generadas por ambos sentidos no se corresponden. En el capítulo anterior justamente describíamos el efecto contrario —el caso de personas cuya mano no responde de acuerdo con lo que dicta su pensamiento, dando lugar al *síndrome de la mano ajena*. La coherencia de cómo se comporta nuestra mano según nuestras expectativas es lo que nos da la sensación de que nos pertenece, de que es parte de nuestro cuerpo.

Una variación espectacular de este experimento fue desarrollada por el científico sueco Henrik Ehrsson, quien hizo que los sujetos sintieran que se salían de su propio cuerpo, o que su cuerpo era transferido al de un muñeco. Para esto, Ehrsson mostró cómo aplicaba estímulos visuales a un muñeco o cualquier otro avatar, mientras simultáneamente (y fuera de su vista) le aplicaba las mismas estimulaciones (táctiles) a los sujetos. El resultado es que la coherencia de las estimulaciones que veían y que sentían hizo que los sujetos percibieran que físicamente su cuerpo pasaba a ser aquel que Ehrsson estimulaba, aunque fuera el de un muñeco. Más allá de que racionalmente sabían que su cuerpo seguía siendo el mismo de siempre, la ilusión era inevitable[144]. (Usando este principio, en un futuro cercano podrían desarrollarse juegos de realidad virtual en los que el jugador sienta sensaciones táctiles concordantes con lo que acontece en la pantalla, lo cual le daría la sensación de *ser* el avatar del juego y sentir las vicisitudes de este en carne propia.)

Volviendo a *Robocop*, mientras sus miembros mecánicos actúen según su voluntad y expectativas —en el contexto de la discusión del capítulo anterior, generándole la ilusión de libre albedrío— sentirá que son suyos. Lo mismo pasaría con un paciente tetrapléjico. Si los miembros implantados siguen las instrucciones de su cerebro (decodificadas a partir del registro de sus neuronas), es de esperar que el paciente termine sintiendo que los miembros prostéticos son parte

de su cuerpo. En otras palabras, un brazo sintético que no haga absolutamente nada, por más parecido que sea al brazo original, nunca dejará de ser una prótesis externa, mientras que un brazo que medianamente reaccione a la voluntad del paciente, por más distinto que se vea, pasará a ser parte de él.

Pero aun debemos enfrentar un gran dilema. ¿Aceptaríamos amputar los miembros de un paciente tetrapléjico para reemplazarlos por prótesis mecánicas? No tenemos demasiado problema en admitir esto en el contexto de una película de ciencia ficción, pero ¿lo haríamos en la vida real? Pongámonos en el lugar del paciente. Su brazo paralizado es flácido e inútil, pero aún sigue siendo su brazo y, a pesar de no sentir absolutamente nada, no querría verlo lastimado y mucho menos amputado. Si bien ya resolvimos los dilemas sobre la identidad y vimos que el paciente terminará sintiendo que los nuevos miembros son parte de su cuerpo, la decisión no es obvia. Sin embargo, avances más recientes ofrecen soluciones mucho menos dramáticas.


CYBERDYNE

Exoesqueleto HAL (*Hybrid Assisted Limb*) de la firma Cyberdyne (izquierda) y robot ASIMO de Honda (derecha).

Para la inauguración del Mundial de Brasil de 2014, Miguel Nicolelis, uno de los pioneros en neuroprostética y fanático del fútbol[145], prometió algo aparentemente imposible: que un paciente paralítico, frente a todo el estadio y frente a cientos de millones de personas siguiendo la transmisión por televisión, se levantaría de su silla de ruedas, caminaría unos pasos y daría el puntapié inicial usando un exoesqueleto —un soporte robótico externo acoplado a sus miembros— controlado por su cerebro. Finalmente la promesa se fue diluyendo y en la transmisión tan solo se vio una toma de 3 segundos en los que esta persona, sostenida de ambos lados, apenas movió unos pocos centímetros su pierna derecha para “patear” la pelota. Mas la

decepción de lo que iba a ser una demostración extraordinaria contrasta con los avances increíbles en los últimos años. De hecho, hoy día ya se comercializan exoesqueletos que ayudan a ejecutar tareas pesadas y que también colaboran en la rehabilitación de pacientes con accidentes cerebrovasculares —como el modelo HAL de la compañía Cyberdyne. Para que modelos como HAL puedan ser usados por pacientes tetrapléjicos, faltaría entonces controlar los miembros directamente con el cerebro. Pero como viéramos antes, esto ya está dentro de los alcances de la ciencia actual.

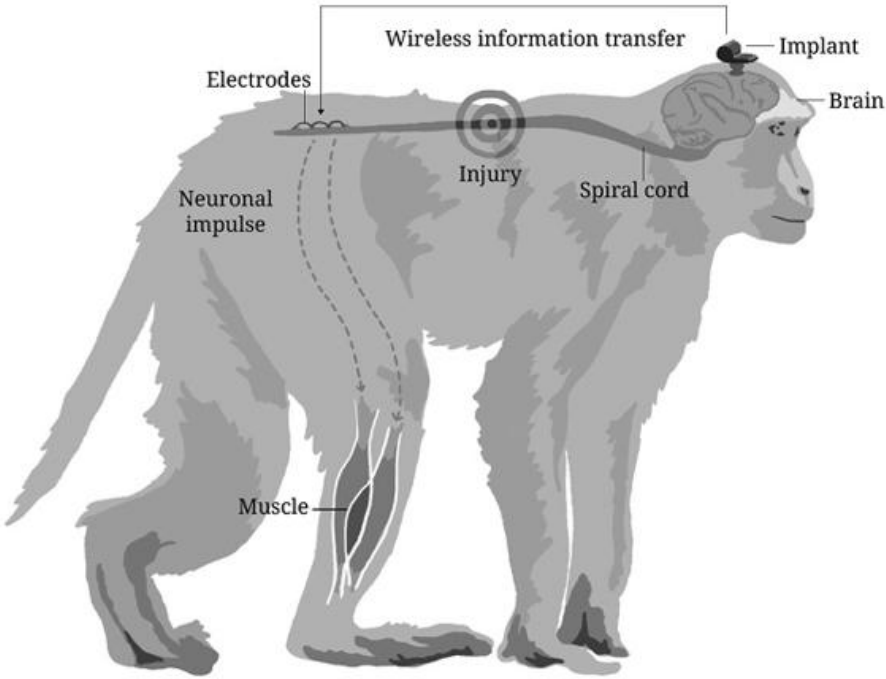
Caminar involucra la activación coordinada de diversos músculos y articulaciones para mantener el balance y un andar fluido. Sin embargo, por más compleja que sea la coordinación de estos movimientos, no hace falta controlar directamente todos estos detalles con el cerebro. Como en el caso del brazo mecánico, de ello se ocupa la robótica. Robots como el ASIMO de Honda ya pueden caminar, correr, saltar y realizar movimientos con una increíble precisión y fluidez. Las instrucciones que manda el cerebro para comandar los miembros prostéticos pueden, entonces, ser muy generales: sentarse, caminar, doblar, correr, etcétera. La implementación de estas instrucciones es un problema de robótica que ya está mayormente resuelto (lo que no quita que no pueda seguir siendo optimizado).

Como en el caso del brazo mecánico, el hecho de no tener que controlar los detalles de los distintos movimientos simplifica muchísimo el problema, y el uso de exoesqueletos nos libra de tener que plantearnos si se justifica o no amputar miembros paralizados en pacientes tetrapléjicos para reemplazarlos por prótesis mecánicas. Pero si aceptamos resignar darles facultades extraordinarias a los pacientes —la fuerza y velocidad de Steve Austin (lo cual no es resignar mucho, ya que no era el objetivo primario, amén de que los pacientes probablemente no quieran verse cargando exoesqueletos como Robocop)— entonces existe una solución mucho más práctica y natural.

Gregoire Courtine es uno de tantos físicos dedicados a la neurociencia, que en su laboratorio en la Universidad de Laussane ha venido explorando la posibilidad de restaurar la locomoción a partir

de estimular eléctricamente la espina dorsal. La gran diferencia con los trabajos que discutiéramos antes es que, en este caso, la idea no es mandar las señales cerebrales a miembros robóticos externos o exoesqueletos, sino a los propios nervios y músculos del paciente. En uno de estos trabajos, Courtine y su equipo realizaron una lesión en la espina dorsal de un mono que le dejó totalmente paralizada la pierna derecha. Luego registraron la actividad en la corteza motora responsable del movimiento de este miembro, decodificaron la información y la transmitieron a los músculos a partir de estimular eléctricamente la espina dorsal (en un punto por debajo de la lesión). Simplificando, al no poder transmitirse la señal neuronal a través de la espina dorsal, lo que hicieron fue leer la información por encima de la lesión (en la corteza cerebral) y transmitirla a un punto por debajo de esta (en la espina dorsal), con lo cual lograron restablecer el andar del mono[146]. La demostración fue espectacular: sin estimulación, el mono cojeaba arrastrando su pata en una cinta giratoria, pero a partir del momento en que estimulaban, comenzaba a mover la pata, adquiriendo un andar mucho más fluido.

Los desarrollos en humanos no son menos espectaculares. A partir de estimular continuamente la espina dorsal, ya es posible que un paciente tetrapléjico pueda volver a mover sus miembros y hasta incluso dejar su silla de ruedas y caminar con un andador, reflotando la promesa de Miguel Nicolelis en el Mundial de Brasil[147].



Idea de una neuroprótesis natural desarrollada en monos. La señal medida a través de un implante en la corteza motora se transmite a electrodos que estimulan la espina dorsal (por debajo de la lesión), que a su vez produce la activación de los músculos de la pierna paralizada. Adaptado de Jackson (2016)[148].

La solución implementada por Gregoire Courtine y sus colegas es muy natural porque reutiliza los propios músculos del paciente para llevar a cabo los movimientos. La desventaja de no usar prótesis mecánicas o exoesqueletos es que ya no podemos contar con los avances de la robótica para resolver la compleja coordinación muscular que permite caminar. Sin embargo, es esperable que la plasticidad neuronal descripta, junto al uso de avanzados algoritmos de decodificación, ayude a realizar estos movimientos; es decir, a pesar de que la conexión entre el cerebro y los músculos varíe (nunca podrá restablecerse la conexión tal cual era), la combinación del uso de algoritmos sofisticados junto con la capacidad de las neuronas de adaptar sus respuestas ayuda a resolver el problema.

Un principio análogo también fue usado para restablecer el movimiento del brazo paralizado de un mono, pero en este caso

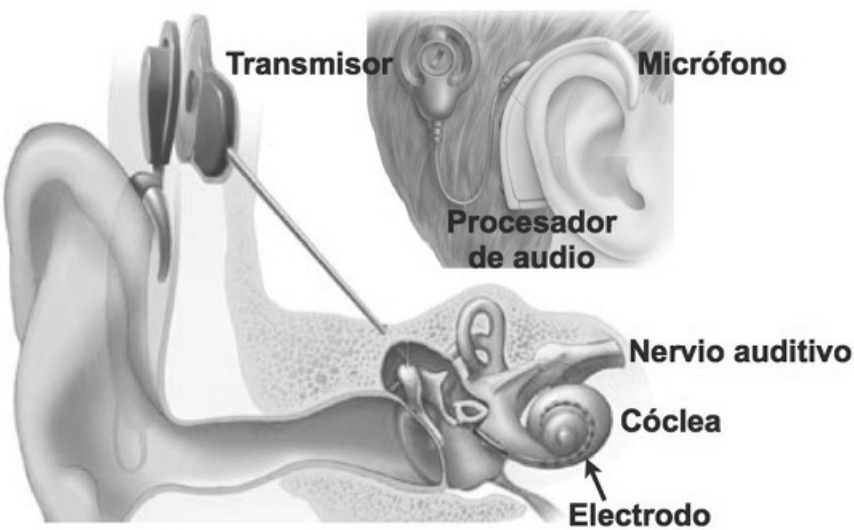
estimulando directamente los músculos del brazo (en vez de la espina dorsal), a partir de la actividad neuronal registrada en la corteza motora[149]. Como en los experimentos descriptos más arriba, este resultado fue seguido pocos años después por resultados similares en humanos, logrando que al estimular los músculos de la mano en base a la señal registrada en el cerebro, un paciente con el brazo paralizado pudiera mover los dedos de su mano y hasta manipular una botella sirviendo su contenido en un vaso[150]. Pero obviamente no es lo mismo agarrar una botella con agua que, por ejemplo, un huevo. Si apretamos el huevo con la misma presión que usamos para sujetar la botella, lo rompemos. Análogamente, si agarramos la botella con la misma presión con la que sostenemos el huevo, se nos cae de la mano. Lo que hace falta entonces es controlar la presión que ejerce la mano para manipular distintos objetos.

Tan acostumbrados estamos a adaptar la presión que hacemos al manipular objetos, que no somos conscientes de ello. Mayormente regulamos la fuerza que usamos en función del *feedback* táctil (procesado en la corteza somatosensorial), dado por presión de nuestros dedos sobre el objeto en cuestión; basta con tratar de prender un fósforo usando guantes gruesos (como los de jardinería) para entender la importancia del *feedback* táctil en nuestros dedos. Un primer paso en este sentido fue dado por Miguel Nicolelis, logrando que un mono pudiera no solo controlar el movimiento de un cursor en una pantalla a partir de la actividad en su corteza motora, sino también distinguir tres objetos virtuales con diferentes texturas, basado en la estimulación aplicada a la corteza somatosensorial, análoga a la estimulación recibida al tocar un objeto[151].

Vimos hasta ahora que avances recientes permitirán en un futuro cercano restablecer la función de las extremidades en pacientes tetrapléjicos (ya sea usando sus propios miembros o reemplazándolos por prótesis sintéticas). ¿Pero qué tan lejos estamos de reemplazar órganos sensoriales? Una cosa es comandar una pierna o un brazo y otra muy distinta es crear prótesis sensoriales que transmitan información al cerebro y permitan interpretar lo que acontece a nuestro alrededor —todo lo que vemos, sentimos, escuchamos.

El éxito de *El hombre nuclear* dio lugar a la creación de *La mujer biónica*, una mujer a la que le reemplazan ambas piernas, el brazo derecho y el oído derecho tras un accidente en paracaídas. Y justamente el tipo de cirugía que se hace para restablecer la audición, a través de implantes de cóclea, constituye el ejemplo más exitoso de prótesis neuronales.

En el espacio exterior hay más que nada silencio. El sonido, tal como lo conocemos[152], está dado por variaciones de presión en el aire y existe en una fracción infinitésima del universo; es propiedad casi exclusiva de nuestro planeta Tierra y quizás también de algún otro planeta, aún desconocido, que pudiera albergar aire en su atmósfera. El sonido en sí mismo, ya sea la voz de un amigo o el estallido de un trueno, ni siquiera existe en la atmósfera. Es una construcción del cerebro a partir de vibraciones de pequeños filamentos en la cóclea (el oído interno) que transforman las variaciones de presión en impulsos nerviosos. En casos de sordera profunda debidos a problemas en el oído interno, la idea es registrar los sonidos mediante un micrófono y transformarlos en impulsos nerviosos, a partir de estimular eléctricamente con electrodos implantados en la cóclea[153].

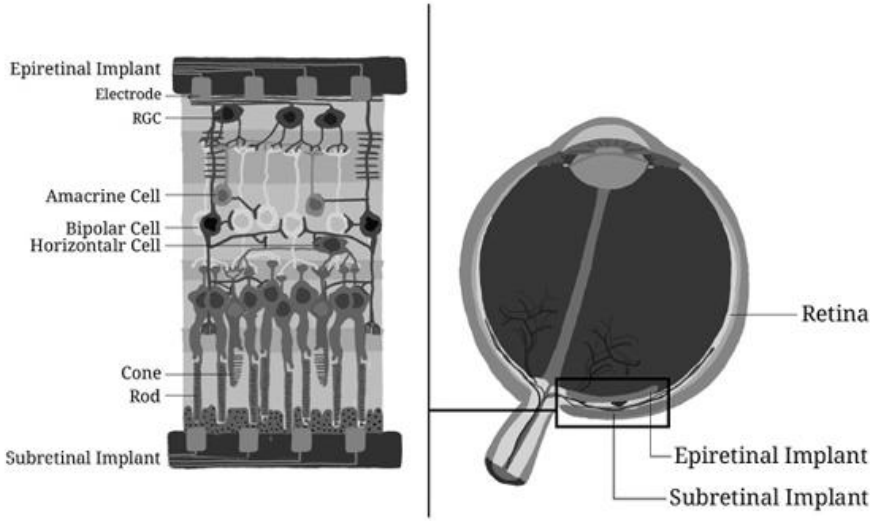


Esquema de implante de cóclea. El sonido captado por el micrófono es procesado y transmitido a un electrodo que estimula eléctricamente la cóclea, generando impulsos nerviosos transmitidos a su vez al cerebro a través del nervio auditivo.

Retomando los ejemplos de las series de ciencia ficción de los 70, el problema de audición de *La mujer biónica* ya está resuelto y hoy día cientos de miles de pacientes pueden escuchar gracias a implantes de cóclea. Mucho más difícil, sin embargo, es lograr que alguien como Steve Austin recupere la visión.

A partir del éxito de los implantes de cóclea, emprendedores y científicos comenzaron a desarrollar implantes de retina. El principio es el mismo, solo que, en vez de captar sonidos, el implante de retina capta los rayos de luz que llegan a los ojos a partir de una lámina con fotorreceptores sobre la retina (o de una cámara de video montada en anteojos especiales) y los transforma en descargas eléctricas aplicadas a través de microelectrodos que estimulan el nervio óptico.

Varios ensayos clínicos han mostrado cómo los implantes de retina hacen que pacientes con ceguera, debido a enfermedades degenerativas de la retina, puedan detectar movimientos, discriminar formas simples (como un cuadrado o un triángulo) y desplazarse en una habitación sin chocar con objetos[154]. Tales avances son muy promisorios, pero aún falta lograr que los pacientes puedan leer o reconocer caras y objetos más sofisticados, para que los implantes de retina pasen a ser usados masivamente. El gran problema es que hay un amplísimo repertorio de cosas que reconocemos con nuestra vista, las cuales dan lugar a distintos patrones de descargas neuronales en el nervio óptico. En otras palabras, hace falta generar patrones de estimulación lo suficientemente específicos y diferenciables en la retina para poder distinguir los objetos frente a nosotros. Obviamente, la cantidad de patrones que pueden identificarse depende del número de microelectrodos que se implanten (a mayor cantidad de electrodos, mayor resolución y capacidad de discriminación). En este sentido, los avances notables en el desarrollo de electrodos miniaturizados en los últimos años llevan a prever grandes progresos en el uso de implantes de retina en un futuro cercano[155].



Esquema del implante de retina. La estimulación a través de electrodos puede hacerse en la capa en donde se encuentran los fotorreceptores de la retina (implante subretinal) o directamente en la capa de neuronas RGC (*Retinal Ganglion Cells*) que constituyen el nervio óptico (implante epiretinal).

Al igual que el sonido, el color no existe a nuestro alrededor; lo que existe son ondas electromagnéticas, rayos de luz que impactan nuestra retina. El color, las formas de todo lo que vemos, son construcciones del cerebro basadas en interpretar impulsos nerviosos generados en la retina. Pero si en el fondo la visión (así como la audición y los otros sentidos) no es más que la interpretación de impulsos nerviosos, nada impide que las ondas electromagnéticas sean captadas y transmitidas al cerebro de otra manera.

A fines de la década del 60, Paul Bach-y-Rita desarrolló un sistema mediante el cual las imágenes registradas por una cámara generaban un patrón de estimulación en la espalda de ciegos[156]. A partir de este sistema, los participantes pudieron reconocer distintos objetos comunes (un teléfono, una silla, una taza, etcétera) y letras mayúsculas, entre otras cosas. Sin embargo, la espalda no posee una gran densidad de receptores y desarrollos posteriores dieron lugar al sistema BrainPort, que usa patrones de estimulación en la lengua, a partir de la señal registrada por una cámara montada en anteojos

especiales. (La ventaja de usar la lengua es que es uno de los órganos con mayor cantidad de terminaciones nerviosas en el cuerpo humano.) Y así como BrainPort convierte información visual en estímulos táctiles, sistemas como vOICe convierten estímulos visuales en sonidos que permiten discriminar distintos objetos. En conclusión, a partir de registrar las ondas electromagnéticas que componen los estímulos visuales, pueden implementarse diversos métodos para transmitir la información al cerebro y lograr que la persona llegue a ver, pero, por supuesto, siempre y cuando el cerebro pueda interpretar esta información.

Analizando un problema planteado por su amigo William Molyneux, John Locke se preguntó por la percepción que tendría un ciego de nacimiento al ver por primera vez (por ejemplo, al observar una esfera y un cubo). Locke concluyó (*Ensayo sobre el entendimiento humano*, 1690, libro II, capítulo 9, sección 8):

> [[I] am of opinion that the blind man, at first sight, would not be able with certainty to say which was the globe, which the cube, whilst he only saw them; though he could unerringly name them by his touch, and certainly distinguish them by the difference of their figures felt.]

> Soy de la opinión que el ciego, al ver por primera vez, no podría decir con certeza cuál es la esfera y cuál es el cubo con sólo verlos; aunque podría inequívocamente nombrarlos a partir de su tacto, y ciertamente distinguirlos a partir de la diferencia sentida al tocarlos.

Argumentos similares, negando la posibilidad de conocimiento innato y resaltando la importancia de la experiencia, fueron esbozados por el obispo George Berkeley en su *Ensayo de una nueva teoría de visión* (1709) y por Hermann von Helmholtz, quien afirmara que “los recuerdos de experiencias previas tienen un rol preponderante e influyente en nuestras observaciones visuales”[157]. No es sorprendente entonces que las personas que comienzan a usar el sistema BrainPort o vOICe, o aquellas sometidas a implantes de retina, necesiten de bastante entrenamiento para empezar a reconocer objetos. No alcanza con conectar un órgano sensorial al cerebro, pues el cerebro tiene que

aprender a interpretar los nuevos estímulos. Aunque la plasticidad neuronal de la que habláramos anteriormente hace que esto sea eventualmente posible.

Sorprendentemente no solo podemos sustituir, sino también aumentar nuestras capacidades sensoriales. Y no me refiero a tener visión infrarroja como las serpientes (o el extraterrestre de la película *Predator*) o a escuchar altas frecuencias como los perros; me refiero, en cambio, a la posibilidad de desarrollar nuevos sentidos. Esto no es desconocido en el reino animal. Algunos peces, particularmente en el Amazonas, pueden detectar campos eléctricos (y usan alteraciones del campo que ellos mismos emiten para poder localizar presas o predadores); algunas aves también detectan campos magnéticos, lo que les da una noción del Norte para poder navegar, etcétera. A partir de esto, podríamos entonces plantearnos qué tipo de información del mundo exterior nos gustaría *sentir* directamente en nuestro cuerpo. Y así como vemos distintos colores, podríamos detectar la intensidad de distintas bandas de wifi (por ejemplo, a partir de estimulaciones con diminutos electrodos en nuestro antebrazo) para buscar los puntos de mejor conexión a Internet. Con sensores análogos, también podríamos sentir el devenir de acciones en las que hayamos invertido nuestro dinero; la cantidad de menciones en Twitter que nos involucran, el número de *likes* en nuestras entradas en Facebook, la cantidad de ventas *online* de nuestro nuevo libro mientras lo promocionamos, o el rating de un programa de radio o televisión en el que estamos involucrados[158]. No sería raro que esta sensación directa de nuestra situación económica o nuestra popularidad en los medios termine generando distintas reacciones cerebrales —por ejemplo, descargas de adrenalina o activaciones de áreas relacionadas con la sensación de recompensa— así como naturalmente se generan a partir de las señales provenientes de los sentidos.

Las posibilidades de aumentar nuestros sentidos son increíbles. Pero estoy muy lejos de sugerir algo así. Internet y los *gadgets* de nuestros tiempos ya nos tienen saturados de información noche y día, y el aumento de los sentidos no haría más que exacerbar el problema. Sería como estar permanentemente *online*, como tantos adolescentes

que andan por la vida hipnotizados con las pantallas de sus celulares; sería quitar el foco de lo más importante, de ejercer la capacidad única que tenemos de pensar, abstraer y extraer un significado de las cosas, para asociarlas con otras y ponerlas en contexto. De hecho, se estimó que en nuestros días estamos expuestos a una cantidad de información equivalente a la contenida en ciento setenta y cuatro periódicos por día; cinco veces más que en la década del 80[159]. Tal bombardeo de información atenta contra aquello que nos hace humanos y que nos distingue de otros animales o las computadoras. Este es un tema harto interesante y de muchísima actualidad, pero cierro aquí la discusión sobre el exceso de información de nuestros días porque ya bastante he despotricado contra ello —particularmente contra el uso de celulares inteligentes y el hecho de estar permanentemente "conectados"— en dos libros anteriores[160].

Vimos entonces que avances de los últimos años permiten vislumbrar la posibilidad muy real y cercana de usar prótesis neuronales para restablecer la función de los distintos miembros y reemplazar o incluso aumentar los sentidos. Lo que era ciencia ficción hasta hace unos años, ya se está experimentando en humanos a partir de desarrollos extraordinarios. La discusión sobre el reemplazo de distintas partes del cuerpo nos llevó a plantearnos el problema de la identidad, aquello que define que una persona sea quien es y no otra. Vimos también que la identidad puede definirse a partir del cuerpo o de la memoria, aunque en realidad esta distinción no tiene mucho sentido, pues implícitamente asume una visión dualista. Después de todo, el cerebro es también parte de nuestro cuerpo y las memorias no son más que la actividad de neuronas conectadas entre sí. Nos queda, sin embargo, discutir la posibilidad de una sustitución mucho más delicada y controversial.

Si nos trasplantaran el corazón de una persona, llamémosla Juan, seguiríamos siendo los mismos pero con un corazón nuevo; si nos trasplantaran su hígado también seguiríamos siendo los mismos pero con otro hígado; lo mismo con el páncreas, los intestinos, los ojos, y así podemos seguir con todos los órganos, pero con una gran excepción: si nos trasplantaran el cerebro de Juan (algo hasta ahora

imposible, pero imaginemos por un momento que pudiera serlo) entonces ya no alegaríamos que seguimos siendo los mismos con un nuevo órgano; diríamos, en cambio, que a Juan le han trasplantado un nuevo cuerpo. Como argumentamos anteriormente, el cerebro es el almacén de la identidad, de la memoria. Mas si en vez de reemplazar nuestro cerebro pudiéramos preservarlo, o tal vez replicarlo, ¿podríamos perpetuarnos más allá de la muerte? Este es la gran pregunta con la que cerraremos este libro, pero antes nos queda hablar sobre la memoria y la influencia que en esta tienen los sueños.

121 Quizás a más de uno que haya crecido en los 70 este texto le traiga muchos recuerdos. En la versión original en inglés el texto es muy distinto, pero dice básicamente lo mismo.

122 Para detalles, ver:

- Miguel Nicolelis. “Action from thoughts”. *Nature* 409: 403-407; 2001.
- John Donoghue. “Bridging the Brain to the World: A Perspective on Neural Interface Systems”. *Neuron* 60: 511-521; 2008.
- Bijan Pesaran, Sam Musallam and Richard Andersen. “Cognitive neural prosthetics”. *Current Biology* 16: R77-R80; 2006.

123 El descubrimiento de Richard fue hecho a través de experimentos con monos. Para una revisión de estos resultados ver: Richard Andersen and Christopher Buneo. “Intentional maps in posterior parietal cortex”. *Annual Review of Neuroscience* 25: 189-220; 2002.

124 No es casual que brinde el ejemplo de Christopher Reeve. Tras su accidente y con una fuerza de voluntad extraordinaria, creó una fundación para apoyar el desarrollo de la neuroprostética. No son pocos los estudios en este campo que estuvieron, al menos en parte, subvencionados por su fundación. Como prueba de su tesón y esperanza, hasta el día de su muerte estaba convencido de que gracias a los avances de la ciencia volvería a caminar. Lamentablemente, debería haber vivido unos años más para ver su sueño hecho realidad.

125 Ver por ejemplo: Rodrigo Quian Quiroga, Larry Snyder, Aaron Batista, He Cui and Richard Andersen. “Movement intention is better predicted than attention in the posterior parietal cortex”. *Journal of Neuroscience* 26: 3615-3620; 2006.

126 En este caso, los registros neuronales fueron hecho en la corteza motora. Para más detalles, ver: Meel Velliste, Sagi Perel, M. Chance Spalding, Andrew Whitford and Andrew Schwartz. “Cortical control of a prosthetic arm for self-feeding”. *Nature*

453: 1098-1101; 2008.

127 Entre otros, un artículo en *The New York Times* (“Monkeys Think, Moving Artificial Arm as Own”) publicado el 29 de mayo de 2008.

128 Para más detalles, ver:

- Leigh Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Masse, John Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney Cash, Patrick van der Smagt and John Donoghue. “Reach and grasp by people with tetraplegia using a neurally controlled robotic arm”. *Nature* 485:372-375; 2012.
- Tyson Aflalo, Spencer Kellis, Christian Klaes, Brian Lee, Ying Shi, Kelsie Pejsa, Kathleen Shanfield, Stephanie Hayes-Jackson, Mindy Aisen, Christi Heck, Charles Liu and Richard Andersen. “Decoding motor imagery from theposterior parietal cortex of a tetraplegic human”. *Science* 348: 906-910; 2015.

129 Por lejos, lo que más desean los pacientes con parálisis no es volver a caminar, sino poder usar sus manos para alcanzar objetos, alimentarse por sí mismos y otras tantas tareas para las cuales constantemente necesitan asistencia. Ver por ejemplo: Kim Anderson. “Targeting recovery: priorities of the spinal cord-injured population”. *Journal of Neurotrauma* 21: 1371-1383; 2004.

130 Para más detalles, ver:

- Leigh Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Masse, John Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney Cash, Patrick van der Smagt and John Donoghue. “Reach and grasp by people with tetraplegia using a neurally controlled robotic arm”. *Nature* 485:372-375; 2012.
- Tyson Aflalo, Spencer Kellis, Christian Klaes, Brian Lee, Ying Shi, Kelsie Pejsa, Kathleen Shanfield, Stephanie Hayes-Jackson, Mindy Aisen, Christi Heck, Charles Liu and Richard Andersen. “Decoding motor imagery from theposterior parietal cortex of a tetraplegic human”. *Science* 348: 906-910; 2015.

131 Esto se debe a varios factores, entre ellos, pequeños movimientos de los electrodos y cambios de impedancias generados por la cicatrización de sangre a su alrededor.

132 Para más detalles, ver: Beata Jarosiewicz, Steven Chase, George Fraser, Meel Velliste, Robert Kass and Andrew Schwartz. “Functional network reorganization during learning in a brain-computer interface paradigm”. *Proceedings of the National Academy of Science of USA* 105: 19486-19491; 2008.

133 En realidad, no todas las neuronas modifican su actividad. Para optimizar resultados, la plasticidad neuronal se complementa con el uso de algoritmos adaptativos, en los cuales se ajustan parámetros para mejorar la performance. Para más detalles, ver:

- Amy Orsborn, Helene Moorman, Simon Overduin, Maryam Shanechi, Dragan Dimitrov and Jose Carmena. “Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control”. *Neuron* 82, 1380–1393; 2014.

- Krishna Shenoy and Jose Carmena. “Combining decoder design and neural adaptation in brain-machine interfaces”. *Neuron* 84: 665-680; 2014.

134 De hecho, los resultados obtenidos en el paciente de la figura de p. 209 fueron obtenidos 5 años después de habérsele implantado los electrodos. Para más detalles, ver: Leigh Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Masse, John Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney Cash, Patrick van der Smagt and John Donoghue. “Reach and grasp by people with tetraplegia using a neurally controlled robotic arm”. *Nature* 485:372-375; 2012.

135 La remake de *Robocop* abunda en guiños filosóficos. El nombre de pila del doctor Norton es una clara referencia a Daniel Dennett, de quien habláramos en capítulos anteriores. En la película el senador *Hubert Dreyfus* se opone al uso de robots para el control de la seguridad, en línea con la visión del filósofo del mismo nombre, quien ha sido un crítico notable del uso de la Inteligencia Artificial.

136 Plutarco; *Vidas paralelas*; Vida de Teseo; XXIII, 1).

137 Sin embargo, en el júbilo de su regreso triunfante, Teseo olvida enarbolar una bandera blanca, como había prometido a su padre, el rey Egeo, en caso de retornar victorioso. Egeo subía todos los días a un promontorio para mirar el horizonte, esperando tener noticias de su hijo. Al ver que la embarcación retorna sin la bandera blanca, asume que Teseo a muerto a manos del Minotauro y se arroja al mar, el cual a partir de esta leyenda se llama mar Egeo.

138 Esto lo relata Platón al comienzo de *Fedón* para explicar la demora en la condena de Sócrates (ver capítulo 1), ya que durante la travesía del barco de Teseo a la isla de Delos estaban prohibidas las ejecuciones.

139 Una de las versiones más conocidas de las paradojas de sorites es la paradoja del pelado. Si a un pelado le agregamos de a un pelo a la vez, tras cada pelo que le agregamos seguiremos diciendo que es pelado, por más pelo que eventualmente termine teniendo al repetir una y otra vez este procedimiento.

La discusión sobre la percepción del cambio estaba muy arraigada en la filosofía de la antigua Grecia, a partir de las posturas contrapuestas de Heráclito, para quien todo cambia, y de Parménides, quien afirmaba que la percepción del cambio es irreal. Aristóteles luego reconciliaría ambas corrientes con su teoría del acto y la potencia (Aristóteles, *Metafísica,* libro IX).

140 John Locke. *An essay concerning human understanding* II. XXVII, 15.

Un razonamiento similar fue propuesto por el filósofo británico Bernard Williams, quien consideró el caso de un científico loco que intercambia nuestra mente con la de otra persona. El científico nos dice que va a torturar a una de las dos personas y nos da a elegir: ¿preferiríamos que torture el cuerpo de la otra persona, pero con nuestra mente, o nuestro cuerpo con la mente de la otra persona? Si elegimos la primera opción, favorecemos la idea de que la identidad va con el cuerpo, mientras que, si optamos por la segunda, creemos que la identidad va con la mente, con la memoria. Bernard Williams. “The self and the future”. *The Philosophical Review* 79: 161-180; 1970.

141 Cabe aclarar que esta afirmación también tiene sus bemoles y no es necesariamente aceptada por todos los filósofos. De hecho, la discusión sobre qué es lo que determina nuestra identidad lleva siglos de debate en la filosofía. Para una descripción más detallada, ver por ejemplo el capítulo 6 de: John Hospers (1956): *An introduction to philosophical analysis,* Londres, Routledge.

142 Matthew Botvinick and Jonathan Cohen. "Rubber hands 'feel' touch that eyes see". *Nature* 391: 756; 1998.

143 Uno de tantos videos en YouTube (https://www.youtube.com/watch?v=RaP0MqvkvUw) muestra la hilarante reacción de los sujetos cuando, tras hacerle sentir que la mano de goma es parte de su cuerpo, la martillan o pinchan con alfileres.

144 Para más detalles, ver:

- Henrik Ehrsson. "The Experimental Induction of Out-of-Body Experiences". *Science* 317: 1048; 2007.
- Ed Yong. "Master of illusion". *Nature* 480: 168-170; 2011.

145 Recuerdo haber conocido a Miguel en un congreso en Los Ángeles. Tras un par de minutos manteniendo una conversación un tanto formal, su actitud cambió notablemente cuando empezamos a hablar de fútbol. Para mi asombro, Miguel conocía los técnicos de algunos equipos del Nacional B en la Argentina y me hizo reír a carcajadas con infinidad de anécdotas de la Copa Roca, que solían jugar Argentina y Brasil. El mayor sinsabor del encuentro fue, sin embargo, el hecho de que en ese entonces solía empezar sus charlas ilustrando la complejidad de la coordinación visuomotora con un gol de tiro libre de Brasil a Argentina.

146 Marco Capogrosso *et al.* "A brain-spine interface alleviating gait deficits after spinal cord injury in primates". *Nature* 539: 284-288; 2016.

147 Los resultados mostrando a un paciente paralítico moviendo sus piernas y caminando con un andador fueron obtenidos por el grupo de Gregoire Courtine. Pero más allá de cuál sea el primer grupo que logra hacer caminar un paciente paralítico a través de una prótesis neuronal —algo que está muy cerca de ser realidad— tomo esto como un avance de la ciencia en general, apoyado en un gran número de resultados por varios grupos, principalmente en los últimos 5 a 10 años. Hasta ahora la estimulación en humanos para restablecer la locomoción ha sido *open loop* (a circuito abierto), usando una estimulación continua para recuperar parte de la movilidad de los miembros. Ver: Claudia Angeli, Reggie Edgerton, Yury Gerasimenko and Susan Harkema. "Altering spinal cord excitability enables voluntary movements after chronic complete paralysis in humans". *Brain* 137; 1394-1409; 2014.

Lo que aún falta es cerrar el circuito y hacer que la estimulación se efectúe a partir de la actividad de las neuronas en la corteza motora, como ya se hizo con monos.

148 Andrew Jackson. "Spinal-cord injury: Neural interfaces take another step forward". *Nature* 539: 177-178; 2016.

149 C. Ethier, E. Oby, M. Bauman and L. Miller. "Restoration of grasp following

paralysis through brain-controlled stimulation of muscles". *Nature* 485: 368-371; 2012.

150 Chad Bouton *et al.* "Restoring cortical control of functional movement in a human with quadriplegia". *Nature* 533: 247-250; 2016.

151 Joseph O'Doherty, Mikhail Lebedev, Peter Ifft, Katie Zhuang, Solaiman Shokur, Hannes Bleuler and Miguel Nicolelis. "Active tactile exploration using a brain-machine-brain interface". *Nature* 479: 228-231; 2011.

152 Menciono "el sonido tal como lo conocemos" refiriéndome al sonido producido por variaciones de presión en el aire. Por supuesto, los sonidos pueden también transmitirse a través de sólidos, líquidos y gases.

153 Cabe aclarar que el implante de cóclea es muy distinto a un audífono, el cual no hace más que amplificar el sonido registrado por un micrófono. El implante de cóclea, en cambio, transforma el sonido en impulsos eléctricos. Para más detalles, ver: J. Rauschecker and R. Shannon. "Sending sound to the brain". *Science* 295: 1025-1029; 2002.

154 Para más detalles, ver:

- J. Mills, A Jalil and P. Stanga. "Electronic retinal implants and artificial vision: journey and present". *Eye* 31: 1383-1398; 2017.
- J. Weiland and Marck Humayn. "Retinal prosthesis". *IEEE Transactions in Biomedical Engineering* 61: 1412-1424; 2014.
- E. Zrenner. "Will retinal implants restore vision?" *Science* 295: 1022-1025; 2002.

155 Ver, por ejemplo: James Jun *et al.* "Fully integrated silicon probes for high-density recording of neural activity". *Nature* 551-232-236; 2017.

156 Paul Bach-y-Rita, Carter Collins, Frank Saunders, Benjamin White and Lawrence Scadden. "Vision substitution by tactile image projection". *Nature* 221: 963-964; 1969.

157 Hermann von Helmholtz [1878]: "The facts of perception", en *Selected writings of Hermann von Helmholtz,* Middletown, Wesleyan University Press; 1971.

Para una descripción más detallada de este tema, ver el capítulo 3 de *Qué es la memoria*, op. cit.

158 Algunas de estas ideas fueron presentadas en una muy entretenida charla TED de David Eagleman en 2015.

159 M. Hilbert y P. Lopez. "The world's technological capacity to store, communicate and compute information". *Science*, 332: 60-65; 2011.

160 *Borges y la memoria*, op. cit.; *Qué es la memoria*, op. cit..

VIII. Inception

Construcción del sueño

Probablemente *Inception* (2010) sea la obra maestra de Christopher Nolan, y no es poco decir del director de la trilogía de *Batman*, *Memento*, *The Prestige* y *Dunkerke*.

Un grupo de especialistas liderados por Dom Cobb se dedican a hacer espionaje industrial, robando secretos del subconsciente de personas durante el sueño. La idea es que al estar despiertos contamos con mecanismos de defensa para proteger nuestros secretos, mientras que en el sueño estas defensas bajan y los pensamientos son susceptibles de ser robados. Cobb y sus secuaces se encargan entonces de generar "sueños compartidos", en los cuales pueden introducirse para así inducir a la víctima a revelar el secreto que buscan. Mas para que la maniobra no sea descubierta, la extracción tiene que ser muy sutil y enmascarada en distintos niveles de sueños. De esta manera, la trama nos sumerge en el fascinante mundo de los sueños, los sueños dentro de los sueños, y cómo estos se articulan e interactúan entre sí: por ejemplo, la caída al vacío de un auto en un nivel de sueño genera una momentánea falta de gravedad en el sueño dentro de este y una avalancha de nieve en otro aún más profundo.

La película combina cinco realidades —cada una en sucesivos niveles de sueño, con un lugar y un tiempo propios— y abunda en reflexiones sobre lo que experimentamos al dormir. En el sueño se violan las leyes de la física y los tiempos se alargan, lo cual nos puede llevar a estar durante horas cayendo en el vacío sin llegar nunca al piso; el mundo de los sueños es inestable y carece de una coherencia narrativa; asimismo, tanto el sueño como la vigilia son una construcción del cerebro, pero en el sueño esta construcción se da exclusivamente a través de la imaginación, mientras que en la vigilia la construcción de nuestra realidad se da a partir de estímulos exteriores.

En varias entrevistas, Nolan se confiesa un admirador de Borges y

resalta su influencia en la elaboración de la trama. Destaca particularmente “El milagro secreto” (*Ficciones*, op. cit.), en donde un escritor de ascendencia judía ruega a Dios que le otorgue un año más de vida para terminar su obra cumbre antes de ser fusilado por el ejército nazi tras la invasión de Praga. Dios accede, pero de una manera muy peculiar: detiene un año el tiempo físico justo antes de ser fusilado y allí, en el paredón, enfrentando al pelotón que le apunta, el escritor repasa y corrige mentalmente la trama de su obra, hasta que finalmente encuentra las palabras para el último epíteto, momento en el cual el tiempo retoma su curso y es fusilado. Un dilema parecido se plantea en *Inception*, en donde uno de los conspiradores es herido de muerte y sus secuaces deciden llevarlo a un nivel de sueño más profundo para prolongarle la vida, ya que allí el tiempo transcurre más lentamente[161].

Otra clara influencia de Borges es “Las ruinas circulares” (*Ficciones*, op. cit.). Como los “arquitectos” que diseñan los sueños en la película de Nolan, el cuento de Borges describe a un personaje que construye a una persona en sus sueños. La monumental tarea, imaginando cada órgano, cada rasgo y hasta cada uno de sus pelos, le lleva años. Mas tras haber logrado su cometido, en un gran incendio nota que el fuego no lo quema y se da cuenta que él mismo está siendo soñado por otro.

Borges reiteradamente juega con la aparente dificultad de distinguir entre la realidad y el sueño; una pregunta que lleva siglos de discusión en la filosofía. De hecho, una de sus antologías (*El libro de sueños*) incluye un famoso cuento del antiguo filósofo chino Zhuang Zhou, quien ya se planteaba este problema hace más de dos mil años: “Un día soñé que era una mariposa y al despertar no sabía si era Zhuang Zhou quien había soñado ser una mariposa, o una mariposa que ahora soñaba ser Zhuang Zhou”.

El sueño de la mariposa de Zhuang Zhou, por el artista chino Luo Zhi (siglo XVI).

Así como Borges y Zhuang Zhou, *Inception* nos lleva a preguntarnos por la realidad y cómo sabemos que no estamos soñando. Nolan también alega haberse inspirado en películas como *Matrix*, en donde los humanos viven una realidad ficticia generada por una gran Inteligencia Artificial. *Matrix* desafía la existencia de nuestra realidad, mientras que en *Inception* Nolan invierte la pregunta y se plantea si no podríamos construir una realidad alternativa en la mente de otra persona (en la cual pudiéramos sumergirnos para robarle sus secretos). Mas al discutir sobre la posibilidad de estar soñando tenemos, ante todo, que distinguir dos alternativas: Podemos preguntarnos si *toda nuestra vida* es un sueño, o podemos, en cambio, tomar una postura menos dramática, aceptar que la realidad existe, y

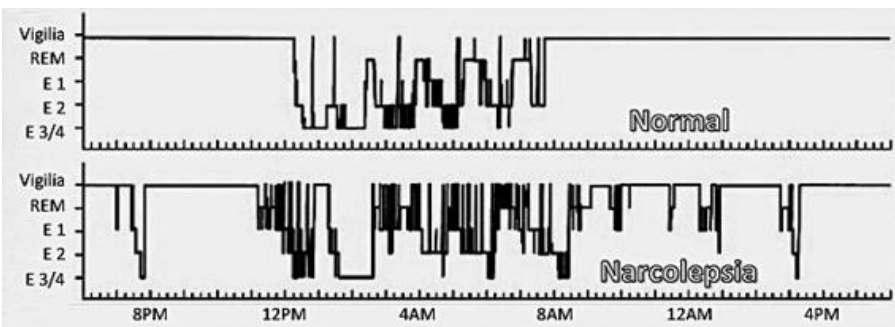
cuestionarnos si *en este momento* no estaremos soñando.

La primera versión, dudando de la realidad, es el planteo que se hiciera Descartes en sus *Meditaciones* (primera Meditación), preguntándose si un artero y poderoso genio maligno podría hacer que "el cielo, el aire, la tierra, los colores, las figuras, los sonidos y las demás cosas exteriores, no son sino ilusiones y ensueños, de los que él se sirve para atrapar mi credulidad". Más cercana en el tiempo, la concepción de Descartes es el argumento del *cerebro en el frasco* que presentáramos en el capítulo 4 al hablar sobre *Matrix*. Sin embargo, a partir de los argumentos de Dennett vimos que una construcción ficticia de la realidad no puede incluir los infinitos detalles de nuestras interacciones con el mundo y tarde o temprano el engaño sería evidente. (De hecho, una de las personas que Cobb busca embaucar en *Inception* se da cuenta de que está en una realidad ficticia, al entender que la alfombra de su oficina no es de lana, como la original, sino de poliéster.) También vimos que Ludwig Wittgenstein mostró que si no existiera la realidad exterior, si no tuviéramos otras personas con quien interactuar, sería imposible el conocimiento.

Aceptando entonces que la realidad existe y que nuestra vida no puede ser solo un sueño, analicemos la otra alternativa: ¿qué evidencia tenemos para saber que *en este momento* no estamos soñando? En otras palabras, ¿cómo podemos distinguir sueño de vigilia? Nolan ofrece una solución ingeniosa: Cobb lleva siempre consigo un tótem único (el trompo de la portada de este capítulo) que solo él conoce en detalle; entonces, ante la duda de si está siendo parte del sueño de otra persona, no tiene más que manipular su tótem para saber si es el mismo de siempre o una copia creada por otro dentro de un sueño[162]. (Por supuesto, esto solo tiene sentido en el contexto de la película, ya que en la práctica no andamos cargando tótems que podamos manipular en nuestros sueños.)

Para el común de la gente la distinción entre estar soñando o despierto parece trivial, casi obvia, pero en algunos casos no es tan así. Cuando dormimos pasamos por distintas fases de sueño, las cuales pueden clasificarse a partir de patrones estereotipados en el electroencefalograma. En una persona normal las fases de sueño se

repiten cíclicamente durante la noche (aproximadamente cada 90 minutos), mientras que en personas con trastornos de sueño la transición entre estas fases es mucho más desordenada. Este es particularmente el caso de pacientes con *narcolepsia*, un desorden del sueño caracterizado por una excesiva somnolencia durante el día, en la que breves fases de sueño aparecen entremezcladas con los períodos de vigilia. La falta de una clara distinción entre sueño y vigilia hace que los pacientes con narcolepsia crean que experiencias vividas en el sueño ocurrieron en la realidad, a veces generando situaciones dramáticas. Los ejemplos abundan: una estudiante recordaba muy vívidamente haber bajado de un autobús y ser recogida por una persona que abusó de ella en un auto; sin embargo, tras varios testimonios y análisis psicológicos, la policía pudo determinar que se había quedado dormida en el autobús y había soñado la situación; una mujer casada estuvo mucho tiempo compungida creyendo haber tenido relaciones extramatrimoniales, algo más tarde comprobaría que no era más que un sueño. Asimismo, a partir de sueños muy vívidos un hombre denunció a su vecino por tirar constantemente basura en el cesto de su casa y luego tuvo problemas con su jefe al creer haber tenido relaciones sexuales con su mujer[163].



Fases de sueño y vigilia en un sujeto normal (arriba) y en una persona con narcolepsia (abajo)[164]. En la persona con narcolepsia las transiciones entre las distintas etapas del sueño (REM y etapas 1, 2, 3 y 4) son mucho más desordenadas que el sujeto normal y también se dan durante la vigilia.

Más allá de los casos de narcolepsia y algunas otras patologías específicas (como la esquizofrenia), en general es muy raro que

alguien confunda sueño con vigilia[165]. Pero ¿qué nos permite hacer tan naturalmente esta distinción?

En su primera Meditación, Descartes argumenta que "no hay indicios concluyentes, ni señales que basten para distinguir con claridad el sueño de la vigilia". Sin embargo, concluye luego que la continuidad de las experiencias es lo que distingue la vigilia del sueño. Dice Descartes (sexta Meditación):

> Pues ahora advierto entre ellos [la vigilia y el sueño] una muy notable diferencia: y es que nuestra memoria no puede nunca enlazar y juntar nuestros sueños unos con otros, ni con el curso de la vida, como sí acostumbra a unir las cosas que nos acaecen estando despiertos. […]
> […] cuando percibo cosas, sabiendo distintamente el lugar del que vienen y aquel en que están, así como el tiempo en el que se me aparecen, y pudiendo enlazar sin interrupción la sensación que de ellas tengo con el restante curso de mi vida, entonces estoy seguro de que las percibo despierto, y no dormido.

Thomas Hobbes, contemporáneo de Descartes, fue muy crítico de este último argumento. Hobbes razonó que en principio uno podría también soñar que lo que acontece en un momento dado está relacionado con los eventos pasados y, por lo tanto, el principio de continuidad no distinguiría sueño de vigilia (Hobbes, "Terceras objeciones"). Sin embargo, más allá de que la objeción de Hobbes sea teóricamente posible, el sueño carece de continuidad temporal. En mis sueños puedo estar en mi oficina escribiendo este libro y luego montar un elefante que me lleve a recorrer la India. La falta de estímulos exteriores me hace perder la coherencia de la realidad, mientras que cuando estoy despierto, el tiempo progresa linealmente y no puedo ir de un lugar a otro sin haber viajado por los puntos intermedios; lo que me acontece en este momento está intrínsecamente ligado a lo que me acaba de pasar y a lo que me pasará en los próximos minutos.

Estamos seguros de no soñar en la vigilia, pero como argumenta Nolan en *Inception*, nuestros sueños también aparentan ser reales mientras duran y solo notamos sus incoherencias al despertar. El problema es que cuando dormimos nuestro estado de conciencia es muy limitado y no llegamos a preguntarnos si lo que estamos viviendo

es un sueño. Sin embargo, esto no siempre es así.

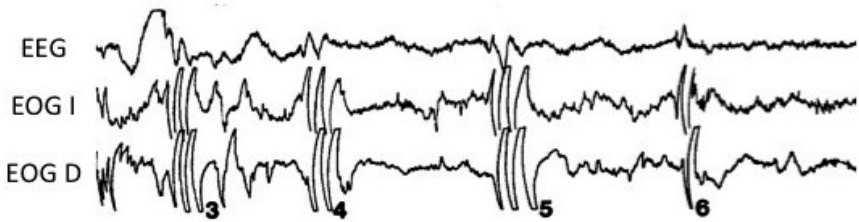
En lo que se conoce como *sueños lúcidos* podemos plantearnos estas preguntas y eventualmente darnos cuenta de que no estamos despiertos. La trama de *Inception* justamente se basa en que los ladrones de secretos son conscientes de estar en un sueño en el cual ejecutan su engaño.

Al compartir por primera vez un sueño con una estudiante que acaba de reclutar en su grupo, Cobb dice que nunca recordamos el comienzo de nuestros sueños; simplemente aparecemos en una situación y en un lugar sin saber cómo. En seguida, tomando un café en París, le pregunta a la estudiante: ¿Cómo terminamos en este restaurante? La joven se muestra confundida, duda un instante y finalmente se da cuenta de que está soñando. Este tipo de preguntas son las que en principio podemos hacernos al dormir para testear la realidad. Por ejemplo, podemos preguntarnos cómo llegamos a un determinado lugar; podemos desafiar las leyes de la física y tratar de atravesar objetos con la mano, o mantenernos suspendidos en el aire más de la cuenta al saltar; podemos taparnos la nariz y tratar de respirar con la boca cerrada; podemos leer algunas líneas de un texto y ver si tienen sentido[166]. A partir de estas pruebas podríamos verificar si estamos teniendo un sueño lúcido.

Alrededor de la mitad de las personas afirman haber tenido al menos un sueño lúcido en su vida y aproximadamente un tercio de estas tienen más de un sueño lúcido por mes[167]. Pero la idea de los sueños lúcidos no es tan fácil de aceptar; después de todo, el sueño es generalmente caracterizado como un estado en donde perdemos la consciencia. Entonces, más allá de reportes subjetivos, ¿qué evidencia científica tenemos para afirmar que una persona es consciente de estar soñando? La búsqueda de tal evidencia fue la tarea que se propuso el psicólogo Stephen LaBerge durante su doctorado en la Universidad de Stanford; algo aparentemente imposible, pues el sentido común nos dice que en principio no podemos comunicarnos con una persona durante el sueño, para que eventualmente reporte ser consciente de lo que está pasando.

En el sueño realizamos movimientos oculares, los cuales producen

variaciones del campo eléctrico que pueden ser captadas con un "electrooculograma" (EOG)[168]. Usando este principio, LaBerge pidió a sujetos que alegaban tener sueños lúcidos que hicieran movimientos oculares preestablecidos (mover los ojos de izquierda a derecha), para así dar testimonio de ser conscientes durante el sueño, pues una vez que notaban los movimientos de ojos en el EOG, los investigadores no tenían más que despertar a la persona y corroborar que las señales registradas en el EOG se correspondían con los movimientos oculares que la persona hizo mientras soñaba.



Registro electroencefalográfico (arriba) y electrooculográfico (de movimientos del ojo izquierdo, en el medio, y del ojo derecho, abajo) de aproximadamente un minuto de duración en un sujeto que realizó varios movimientos preestablecidos (marcados 3 - 4 - 5 - 6), reportando que estaba lúcido durante el sueño[169].

Volviendo al argumento de *Inception*, la clave para poder extraer secretos de las víctimas se basa en la posibilidad de compartir sueños y así acceder a su subconsciente. Vimos en el capítulo 5 que hasta cierto punto leer los sueños de una persona ya es una realidad: tres años después de haberse estrenado *Inception*, el grupo de Yukiyasu Kamitani, en la Universidad de Kyoto, mostró que era posible predecir lo que estaba soñando un sujeto, sobre la base de registros de la actividad cerebral medidos con resonancia magnética por imágenes. Pero ¿será posible compartir un sueño?

Para lograr que dos personas eventualmente compartan un sueño necesitamos que sus cerebros interactúen y estén comunicados. O sea, hace falta que los pensamientos de una persona influyan directamente en los de otra mediante lo que se llama *brain-to-brain communication* (comunicación cerebro a cerebro). Un experimento de este tipo fue realizado en el laboratorio de Miguel Nicolelis (de quien hablamos en

el capítulo anterior). Una rata fue entrenada para reconocer dos estímulos que le indicaban qué palanca debía presionar para obtener su recompensa (una gota de jugo), mientras su actividad cerebral era registrada y transmitida a una segunda rata (estimulando su corteza cerebral) que se hallaba en otra jaula y no veía los estímulos, pero que podía decidir correctamente la palanca a accionar en función de la actividad cerebral que recibía de la primera[170]. Sin embargo, cabe aclarar que hay una diferencia sideral entre transmitir una decisión binaria (la palanca a elegir) y transmitir los contenidos de un sueño. Para que esto último fuera posible, haría falta estimular selectivamente grupos de neuronas en un sujeto a partir de las activaciones registradas en el otro. Por ejemplo, si un sujeto sueña con un cielo color fucsia, habría que interpretar el contenido de este sueño y luego identificar y estimular las neuronas que representan la idea del cielo y del color fucsia en el otro. Esto ya es extremadamente complejo y apenas estaríamos transmitiendo el color del cielo, ni qué hablar de los contenidos específicos de un sueño.

Dejando entonces de lado la posibilidad de manipular los sueños de otra persona, algo que hasta ahora continúa siendo ciencia ficción, nos planteamos algo mucho más tangible pero no menos espectacular: ¿podremos manipular o incluso construir nuestros propios sueños?

Nos hemos referido a los sueños lúcidos. Hay infinidad de cursos, libros y páginas de Internet dedicados a este tema. El motivo es muy sencillo: si somos conscientes de estar soñando, podríamos, por ejemplo, evitar una pesadilla, pues al saber que una situación no es real, se desvanece el miedo que eventualmente puede generarnos. Pero mucho más interesante es el hecho de que, siendo conscientes de nuestros sueños, seríamos capaces tomar decisiones para manipularlos. Por ejemplo, podemos elegir no entrar a una casa lúgubre en donde sabemos que nos esperan infinidad de problemas y decidir, en cambio, dar un gran salto y flotar mágicamente en el aire (como solemos hacer en el sueño). Si pudiéramos conquistar nuestros sueños, las posibilidades serían infinitas: podríamos viajar, encontrarnos con gente que deseamos conocer, volar, ir a otros planetas, llegar al fondo del océano. La posibilidad de influenciar

aquello que acontece durante el sueño es sin lugar a dudas muy atractiva. Pero ¿es realmente posible construir a medida nuestros sueños como en *Inception*? ¿Podremos elegir qué soñar antes de dormir y encontrarnos con nuestra actriz favorita, algún ser querido, o dar un paseo por la Luna?

Esto es lo que se conoce como *incubación de sueños*. Generalmente tendemos a soñar acerca de las cosas que nos suceden durante en el día. Por lo tanto, la incubación de sueños consiste en pensar una y otra vez sobre aquello que queremos soñar antes de dormir. Deirdre Barrett, psicóloga de la Universidad de Harvard, testeó esta hipótesis pidiéndole a un grupo de estudiantes que durante una semana trataran de "incubar" sueños a partir de algún problema que fuera importante para ellos. Luego los estudiantes escribieron lo que recordaban de sus sueños y aproximadamente la mitad de estos estaban relacionados con el tema elegido[171]. El resultado de Barrett no es tan sorprendente considerando la experiencia que generalmente tenemos de continuar pensando sobre los problemas que nos preocupan en la vigilia (a veces encontrando soluciones creativas y novedosas). Pero una cosa es influenciar el contenido de los sueños y otra muy distinta es controlar lo que va a acontecer en ellos.

Los sueños tienen su propia dinámica, la cual no podemos decidir a voluntad. Por más fácil que pueda resultarnos tener sueños lúcidos o incubar sueños, a lo sumo manipularemos el tema general de nuestro sueño, pero no su trama específica. Aunque, por supuesto, determinar el tema de un sueño no es poco, pues prefiero soñar algo que me agrade a terminar en una pesadilla.

Pasamos un tercio de nuestra vida durmiendo, ni más ni menos que 25 años en una persona de 75. Soñamos que caemos, que somos perseguidos, que no pasamos un examen, que estamos desnudos en público, que no podemos correr o que se nos caen dientes. Pero ¿por qué?

En la Antigüedad se creía que los sueños eran presagios. El Antiguo Testamento, y particularmente el Génesis, describe un gran número de sueños en los que Dios se comunica con los hombres. La pintura de la página siguiente representa el sueño de Jacob (Génesis 28, 12):

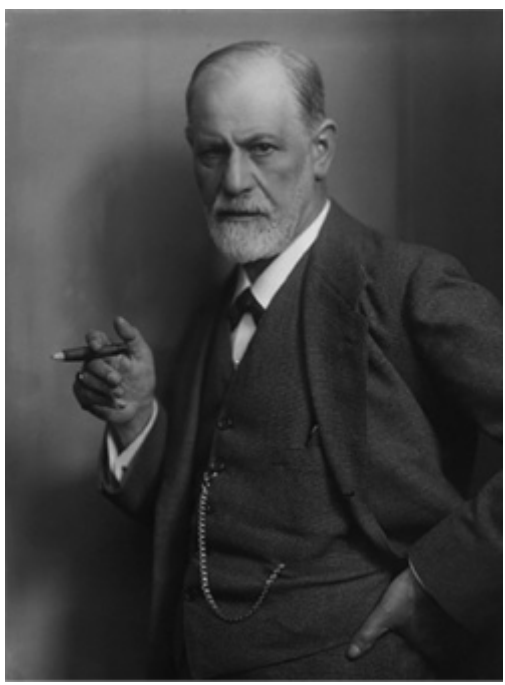
temiendo la venganza de su hermano, Jacob decide irse de su casa; en el camino, haciendo noche en un descampado, sueña que desde allí se levanta una escalera que llega hasta el cielo por donde suben y bajan los ángeles; desde las alturas, Dios le manifiesta que en esa tierra lo protegerá a él y a toda su descendencia, y según las escrituras, al despertar Jacob decide radicarse en ese mismo lugar, al que llamará Bethel.

El sueño de Jacob. Óleo de José de Ribera (1639).

Más allá de esas historias, no existe evidencia científica alguna que avale el rol premonitorio de los sueños y, aun suponiendo que tomemos por ciertos los muy pocos testimonios sobre sueños anticipando el futuro, estos son tan escasos que es razonable atribuirlos a coincidencias al azar. Mucho más lógico es argumentar que dormir es el descanso que necesita el cuerpo, y particularmente el cerebro, tras cada jornada. Pero esto tampoco justifica los sueños. De hecho, probablemente descansaríamos mucho mejor si tuviéramos nuestra mente totalmente en blanco al dormir y no habría motivo alguno para crear los mundos extraordinarios que apenas recordamos al despertar.

El estudio sistemático y riguroso de los sueños comienza con Sigmund Freud (1856-1939), quien sentaría las bases del psicoanálisis. En su obra más famosa, publicada a inicios el siglo XIX, dice que "la interpretación de los sueños es el camino real hacia el conocimiento de las actividades inconscientes de la mente" (*La interpretación de los sueños*, 1900, capítulo VII E). Freud argumenta que los sueños no tienen causas externas (los dioses dándonos a conocer distintas profecías) sino internas —nuestro subconsciente— y de allí el interés en estudiarlos. Para Freud los sueños son como una válvula de escape que da rienda suelta a nuestros deseos reprimidos durante la vigilia (muchas veces de índole sexual), pues al estar despiertos, tenemos mecanismos de censura que nos inhiben de llevar a cabo acciones embarazosas[172].

Los sueños son entonces el momento de liberación de los deseos prohibidos en nuestro subconsciente (aquellos que olvidamos al despertar), aliviando así la frustración de no poder realizarlos cuando estamos despiertos.



Dos pioneros en el estudio de los sueños: Sigmund Freud (izquierda) y Carl Jung (derecha).

En la actualidad, a casi 120 años de publicado su libro, es muy fácil criticar a Freud. Para ponerlo en contexto, cuando él publicaba *La interpretación de los sueños*, Santiago Ramón y Cajal recién estaba dando a conocer su doctrina de la neurona como unidad básica del cerebro, mientras que en nuestros días los avances de la neurociencia permiten medir la actividad de miles de neuronas y ver cómo estas determinan distintos comportamientos.

Freud sostuvo argumentos que hoy sabemos no son ciertos, pero lo importante es que, equivocado o no, generó una discusión que dio lugar a avances notables en nuestro conocimiento. Aunque esto no es excusa para que ciertas ramas de la psicología, y particularmente del psicoanálisis, continúen haciendo un culto de la interpretación de los sueños "a la Freud"[173].

Dejando de lado posibles fallas en la implementación de las ideas de Freud en el psicoanálisis, la crítica a su teoría del sueño es mucho más concreta. Freud plantea que el sueño tiene un efecto liberador, pues si no soñáramos enloqueceríamos al no poder satisfacer nuestros deseos

reprimidos, algo que tiempo después se comprobaría que no es el caso.

A principios de la década del 50, Eugene Aserinsky era un estudiante de fisiología en la Universidad de Chicago, que llevaba adelante su doctorado con Nathaniel Kleitman, quizás el experto en sueño más renombrado de su época. Durante interminables horas, Aserinsky tomaba registros electrooculagráficos (EOG) en bebés, para estudiar los movimientos de los ojos durante el sueño. (Incluso cuando dormimos, los ojos se mueven involuntariamente cada tanto.)

Hasta entonces se creía que el sueño no era más que un período uniforme de descanso, a fin de que el cerebro recupere energía para afrontar un nuevo día. Sin embargo, Aserinsky observaba ahora períodos de sueño en los cuales los ojos paradójicamente se movían como al estar despiertos; de aquí que esta etapa se denomine *sueño paradójico* o REM (del inglés: *Rapid Eye Movement*)[174]. El descubrimiento, inesperado, cambiaría radicalmente la concepción que tenemos sobre el sueño, al cual actualmente clasificamos en REM y no-REM[175]. Los movimientos oculares durante la etapa REM pueden notarse a simple vista (a pesar que la persona tenga los párpados cerrados) y, sorprendentemente, durante miles de años los más grandes pensadores se interesaron por los sueños, pero nunca antes alguien había notado esos movimientos.

La siguiente pregunta que se hizo Aserinsky es obvia. ¿Qué pasa durante el sueño REM? ¿Por qué tenemos esos períodos en los que frenéticamente movemos los ojos de un lado al otro? Hay una posibilidad más que tentadora: en caso de que los sueños se den particularmente durante la fase REM, los movimientos oculares podrían deberse a fijaciones de la mirada explorando las situaciones que vivimos en el sueño. Aserinsky decidió entonces despertar a los sujetos con los que realizaba experimentos mientras dormían y comprobó que la gran mayoría de los sueños efectivamente se daban cuando los despertaba en la fase REM. Hoy día, sin embargo, sabemos que los movimientos oculares no se corresponden con exploraciones visuales durante el sueño y, por otro lado, también sabemos que los sueños se dan igualmente en las fases no-REM, aunque con una probabilidad mucho menor y sin la carga emocional y la bizarra

construcción narrativa del sueño REM[176].

Dimos un gran rodeo contando los pormenores del descubrimiento de la fase REM y su relación con los sueños porque a partir de esa información ahora tenemos todas las herramientas para testear experimentalmente la teoría de Freud. Si el sueño es la liberación de los deseos reprimidos de una persona, privarla de sus sueños —despertándola cada vez que entra en la fase REM— tendría efectos devastadores. Sin embargo, la privación del sueño REM produce ciertos trastornos menores, pero ciertamente no las consecuencias que predijera Freud[177].

Más consistente con estos resultados es la visión de Carl Jung (1875-1961), discípulo y (en un principio) amigo de Freud, quien tomó una posición distinta a la de su mentor sobre los sueños (lo cual terminó haciendo trizas su amistad). Para Jung, en vez de esconder deseos reprimidos, los sueños son una ventana de acceso al inconsciente y tienen una función compensatoria, haciendo que las perspectivas inconscientes complementen las experiencias conscientes de la vigilia. Por otro lado, Jung también argumentó que los sueños proveen visiones de futuros posibles, es decir, no ofrecen premoniciones, como se creía en la Antigüedad, pero nos preparan para afrontar situaciones que podrían llegar a pasar. En otras palabras, si en algún momento enfrentamos una situación novedosa en nuestra vida, es posible que inconscientemente sepamos cómo reaccionar porque habremos experimentado situaciones análogas en el sueño.

Una visión radicalmente distinta sobre los sueños fue propuesta a fines de los años 70 por Allan Hobson y Robert McCarley, dos psiquiatras de la Universidad de Harvard[178]. Para estos investigadores, durante el sueño las activaciones de grupos de neuronas (y aquello que representan) se producen al azar. Por ejemplo, imaginemos que casualmente se activan neuronas que representan los conceptos mar, árbol, búfalo y castillo (activaciones que probablemente estén relacionadas con cosas vividas durante el día). Pero ¿qué tiene que ver esto con los sueños? La idea es que, a partir de estas activaciones, el cerebro construye una narración. Por ejemplo: “Estoy en la playa en donde hay un gran árbol en la arena y

un búfalo que me corre, aunque sin alcanzarme nunca, hasta que finalmente logro perderlo al llegar a un castillo". La historia no tiene sentido, así como tampoco tienen sentido los sueños. Para Hobson y McCarley el hecho de que los sueños sean tan extraños se da justamente porque el cerebro genera un relato basado en activaciones neuronales al azar.

Al hablar de Helmholtz y Bartlett, ya vimos en el capítulo 4 que tanto la percepción como la memoria son construcciones del cerebro. La diferencia con el sueño es que en la vigilia las construcciones se dan a partir de estímulos coherentes y relacionados los unos con los otros porque responden a la realidad (en la vida real no veré un árbol enorme en el medio de la playa, un búfalo y un castillo), pero nuestra percepción de lo que acontece tanto en el sueño como en la vigilia se basa en la capacidad única que tiene el cerebro de construir una realidad a partir de muy poca información. Es algo que inconscientemente hacemos todo el tiempo; tan automático e inevitable que vemos constelaciones representando héroes mitológicos en estrellas distribuidas al azar en el firmamento; tan falaz, que los casinos de Las Vegas muestran las últimas decenas de resultados en la ruleta para que los jugadores ingenuos crean ver un patrón que no existe y pierdan su dinero.

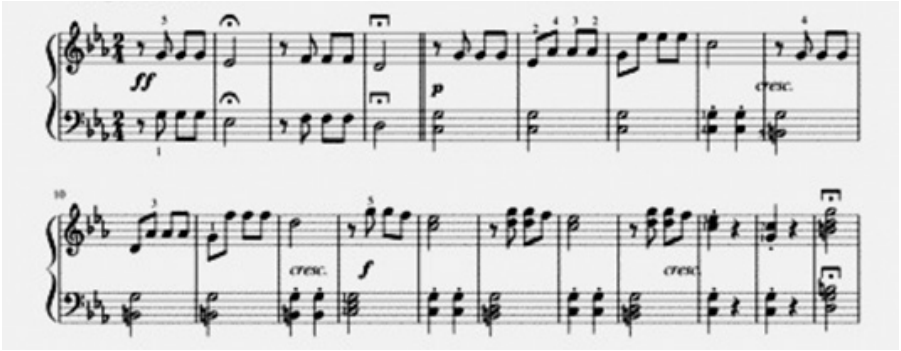
Un proverbio alemán dice *Träume sind Schäume* (los sueños son espuma). En otras palabras, los sueños no son nada, o como se dice más formalmente: son un *epifenómeno*. Un epifenómeno es, por ejemplo, el ruido del motor de un auto: la función del motor es mover el auto, no hacer ruido; el ruido es una consecuencia secundaria. De la misma manera, según la teoría de Hobson y McCarley, el sueño no tiene función alguna y es solo la consecuencia inocua de activaciones al azar en el cerebro.

Pareciera que la neurociencia finalmente le dio el golpe de gracia a la interpretación de sueños de Freud y Jung. Pero a mi entender esta apreciación es algo injusta. Freud y Jung tienen el gran mérito de haber sido pioneros en ver a los sueños como una herramienta de acceso al inconsciente. Más allá de que los sueños no escondan deseos sexuales reprimidos o sean particularmente útiles para compensar el

consciente, sí dan cuenta de nuestras emociones e intereses —aquellas que se manifiestan en los sueños— las cuales son la fuerza motriz que subyace nuestros impulsos y que, por ejemplo, en parte dan lugar a la sensación de libre albedrío y nos llevan a tomar decisiones fundamentadas en un aparente ejercicio de voluntad determinado por nuestro subconsciente.

La visión del sueño como epifenómeno es también problemática, ya que implícitamente esconde una vez más el dualismo cartesiano, la división entre mente y cerebro. Al decir que el sueño es un epifenómeno estamos diciendo que lo que existen son las activaciones de las neuronas, mientras que aquello que representan (y la construcción que hacemos a partir de estas representaciones, es decir, el sueño) no es más que un efecto secundario. Sin embargo, como afirma el materialismo que venimos sosteniendo desde un principio, las activaciones neuronales y aquello que representan son una y la misma cosa. La visión materialista nos lleva nuevamente al problema difícil que describiéramos en el capítulo 2, el cual vale la pena discutir brevemente de nuevo.

Imaginemos, una vez más, que durante el sueño activamos neuronas que representan el mar. No hay agua salada en nuestro cerebro, solo pequeños impulsos eléctricos que son transmitidos de una a otra neurona. Mas los impulsos eléctricos de las neuronas son una representación de la realidad, no la realidad en sí misma, así como la secuencia de ceros y unos en un CD son una representación de distintas piezas de música. Por supuesto, el código es ilegible a simple vista, pero también usamos otras representaciones, como las notas de una partitura.



Primeros compases de la Quinta Sinfonía de Beethoven. Las notas representan una pieza de música, así como la activación de grupos de neuronas representan la realidad exterior en nuestro cerebro.

Una pieza de papel (como una secuencia binaria de números) físicamente no es música, aunque la representa. Podemos, sin embargo, alegar que hay una diferencia fundamental entre una partitura y las activaciones de neuronas en el cerebro. La partitura es en el fondo un "manual de instrucciones" para manipular instrumentos que eventualmente ejecutarán la música. La relación entre la música que escuchamos y la partitura (o la secuencia de ceros y unos que decodificamos en un reproductor de CD) es muy clara. El "problema difícil" de la neurociencia es entender cómo a partir de activaciones de neuronas *sentimos*, por ejemplo, el mar, con sus olas, su olor de algas y el frío intenso al zambullirnos (al menos en las aguas de la costa atlántica argentina). La relación entre las activaciones neuronales y lo que sentimos no es tan obvia como en el caso de la pieza de música. Pero estas sensaciones, lo que denominamos *qualia*, no son más que activaciones de grupos de neuronas representando cosas que están asociadas en nuestro cerebro: al pensar en el mar activamos un grupo de neuronas que representan este concepto, que a su vez activa otros grupos de neuronas representando el olor de algas, la visión de las olas, el frío del agua helada e infinidad de memorias y emociones relacionadas. Por lo tanto, así como la partitura es una representación que permite tocar una pieza de música en un instrumento, la activación de las neuronas corresponde a una representación de algún concepto que, a su vez, da

lugar a la activación de representaciones relacionadas haciéndonos *sentir* las cosas.

Concluimos entonces que no es correcto ver a los sueños como una consecuencia inocua —un epifenómeno— a partir de activaciones del cerebro. Las activaciones neuronales y aquello que representan son la misma cosa, si bien aun nos queda por dilucidar si las activaciones al azar durante los sueños tienen alguna función específica.

Cien años de soledad, del genial Gabriel García Márquez, es tal vez la novela más destacada de la lengua española. Publicada en 1967, marca el punto cumbre del realismo mágico, una corriente literaria desarrollada por escritores latinoamericanos del siglo XX, en donde se combinan magistralmente realidad y fantasía. La imaginación de García Márquez no tiene límite y va desde concebir a una mujer que come tierra y cal rasguñando las paredes, hasta una anciana de casi 150 años que es enterrada en su mecedora; un joven al que siempre lo siguen mariposas amarillas, o un bebé con cola de cerdo al que se llevan las hormigas.

La acción transcurre en Macondo, un pueblo ficticio en donde viven Aureliano Buendía y su numerosa familia. La rutina en Macondo transcurre sin grandes sobresaltos, hasta que un día llega la “peste del insomnio” (¡y qué mejor ejemplo de realismo mágico!). La epidemia se expande rápidamente entre los habitantes, quienes en un principio se alegran de tener más tiempo para terminar todas las cosas que no llegaban a hacer durante el día, aunque poco después descubren el monumental aburrimiento de estar despiertos sin pausa y una consecuencia aún mucho más dramática: el olvido[179]. A medida que se van sucediendo las noches de insomnio, la perdida de la memoria se hace más aparente. Para tratar de sobrellevar la situación, Aureliano Buendía comienza a pegar cartelitos con el nombre de las cosas: yunque, mesa, silla, reloj, puerta. Mas con el creciente avance del olvido, solo el nombre no alcanza y, por ejemplo, cuelga un cartel de la vaca que dice: “Esta es la vaca, hay que ordeñarla todas las mañanas para que produzca leche y a la leche hay que hervirla para mezclarla con el café y hacer café con leche”. José Arcadio Buendía, su padre, decide construir una “máquina de la memoria” con varios

miles de tarjetas que pasan una tras otra en un aparato giratorio y que todas las mañanas le recuerdan los conocimientos adquiridos en su vida. La peste del insomnio hace estragos en Macondo, hasta que un día vuelve el gitano Melquíades trayendo consigo una pócima que cura el olvido[180].

Seguramente alguna vez habremos escuchado que es una muy mala idea quedarse estudiando la noche previa a un examen. Más allá de estar cansados en el examen, el problema es que no habremos asimilado la nueva información a causa de no haber dormido. Para verificar esto, a principios de la década del 20 John Jenkins y Karl Dallenback testearon el rol que tiene el dormir en la memoria, comparando la cantidad de sílabas sin sentido que un grupo de sujetos recordaba a distintos períodos de tiempo (habiendo dormido o no). Como era de esperarse, encontraron que la cantidad de sílabas recordadas disminuía con el tiempo —cuanto más tiempo pasa, menos recordamos—, pero Jenkins y Dallenback también notaron que el desempeño de los sujetos aumentaba notablemente si habían dormido antes de ser evaluados[181].

Queda claro entonces que dormir tiene una función clave en la memoria; pero ¿podemos decir lo mismo de los sueños? O sea, ¿qué es lo que realmente importa: el período de descanso que tenemos al dormir o la construcción misma de los sueños? En el fondo nos preguntamos si los sueños “son solo espuma”, como se desprende del modelo de Hobson y McCarley, o si tienen alguna función, aunque quizás no aquella postulada por Freud o Jung. Para contestar esta pregunta podemos usar el hecho de que los sueños suelen darse en la fase REM, comparando los efectos que tienen tanto la fase REM como la no-REM en la memoria declarativa (la memoria de hechos y vivencias).

Salvo en patologías como la narcolepsia, vimos que las fases del sueño siguen una estructura que se repite aproximadamente cada 90 minutos. El comienzo del sueño está dominado por la fase no-REM y la primera aparición de la fase REM suele darse tras una hora de estar durmiendo. Por lo tanto, una típica siesta corta, de menos de una hora, carece de sueño REM. Un grupo de investigadores alemanes

usaron este principio para evaluar la influencia de la fase no-REM en la consolidación de la memoria. Le dieron a un grupo de sujetos una lista de palabras a recordar y evaluaron cuántas palabras recordaban al poco tiempo, tanto si se mantenían despiertos como si dormían una breve siesta. El resultado sorprendente fue que incluso tras una siesta de tan solo seis minutos los sujetos recordaban más palabras que si permanecían despiertos[182]. En línea con este resultado, otro grupo de investigadores alemanes vio que la consolidación de la memoria era más fuerte si los sujetos solo dormían durante las primeras horas de la noche (desde las 11 a las 2 de la mañana), en la cual domina la fase no-REM, comparado a si dormían más tarde (3 a 6 de la mañana), en la cual domina la fase REM[183].

Estos y otros tantos experimentos muestran claramente que la fase no-REM es vital para la consolidación de la memoria[184]. La evidencia de un rol específico de la fase REM (y particularmente del sueño) en la memoria es, en cambio, relativamente escasa. Para abordar esta pregunta, Robert Stickgold —investigador de la Universidad de Harvard y uno de los expertos más renombrados en el estudio del sueño— evaluó cómo un grupo de sujetos resolvía anagramas o establecía relaciones semánticas entre palabras, inmediatamente después de haber sido despertados en la fase de sueño REM y no-REM[185].

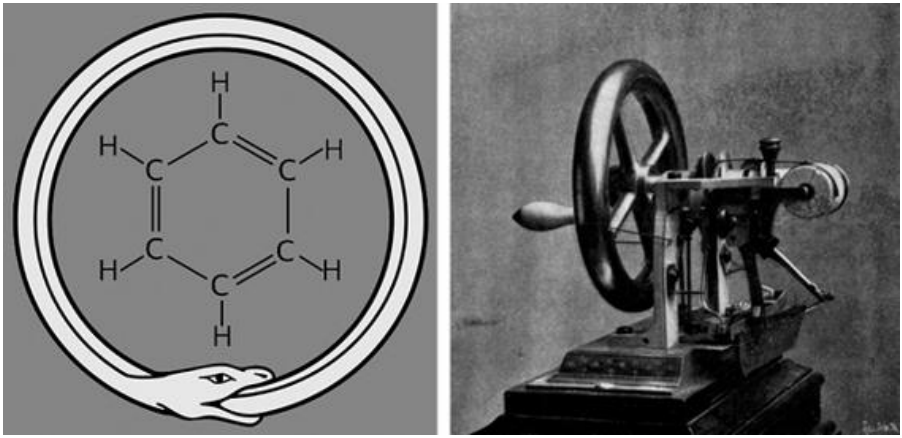
En general, cuando despertamos persiste unos minutos la “modorra” del sueño. Asimismo, al despertar en la fase REM o no-REM el cerebro continúa durante algunos minutos en el estado en que se encontraba hasta despertar del todo. El truco de Stickgold fue realizar los experimentos apenas un par de minutos después de haber despertado a los sujetos, para así asumir que el cerebro continuaba procesando información como en la fase de sueño en la que acababan de estar. En ambos experimentos el resultado concluyente fue que el desempeño de los sujetos era mucho mejor cuando eran despertados en la fase REM, la fase donde se produce la mayoría de los sueños. Stickgold concluyó entonces que la fase REM, y en particular los sueños, favorecen un estado de mayor fluidez de pensamiento comparado con la fase no-REM.

El sueño está influenciado por las situaciones vividas durante el día, aunque dista mucho de ser una repetición de estas memorias. Para mostrar esto, Stickgold y sus colegas preguntaron a un grupo de sujetos por sus sueños y las experiencias de los días anteriores, y observaron que si bien los sueños tocaban los temas generales de estas experiencias (como en los experimentos de Deirdre Barrett que mencionáramos antes) apenas un 1 o 2% de estos las repetían literalmente[186].

Como vimos más arriba, la consolidación de las experiencias en la memoria no se da en el sueño, sino en la fase no-REM, cuando menos soñamos. Stickgold argumenta, a mi entender con mucho criterio, que el sueño en la fase REM tiene una función mucho más interesante: poner estas experiencias en contexto, contrastándolas y asociándolas con el resto de nuestro conocimiento. Cabe notar que las asociaciones elaboradas durante el sueño no están limitadas a las experiencias inmediatas de la vigilia, sino que abarcan toda una gama de conocimiento almacenado. El hecho de probar asociaciones dispares, y no solo aquellas cercanas y coherentes, hace que los sueños sean tan bizarros, pues las activaciones al azar descriptas por Hobson y McCarley testean la relación de nuevas memorias con otras ya adquiridas, generando la maraña de asociaciones que forman nuestra memoria y conocimiento. En conclusión, la fase no-REM afirma las nuevas memorias, mientras que la fase REM y los sueños las ponen en contexto, lo cual provoca no solo que recordemos sino que comprendamos[187].

Las asociaciones que testeamos y establecemos durante los sueños son también la piedra angular de nuestra creatividad. Como vimos al final del capítulo 2, la creatividad se basa en establecer asociaciones entre conceptos dispares; la genialidad de Newton fue el haberse dado cuenta de que la manzana que cae y la Luna que gira a nuestro alrededor responden al mismo fenómeno: la ley de gravedad. Y justamente este tipo de asociaciones dispares son las que suelen darse en el sueño. Por ejemplo, el químico alemán August Kekulé descubrió la estructura cíclica del benceno tras soñar con una serpiente que se mordía la cola. El inventor norteamericano Elias Howe descubrió el

principio de funcionamiento de la máquina de coser tras soñar estar rodeado por un grupo de aborígenes con lanzas. Robert Stevenson imaginó parte de su novela más famosa —*Dr. Jekyll and Mr. Hyde*— durante el sueño, así como Mary Shelley concibió a *Frankenstein* en una pesadilla. Más cercano en el tiempo, Paul McCartney soñó la melodía de "Yesterday", la cual escribió al despertar.



Estructura del benzeno, descubierta por Kekule (1829-1896) tras soñar con una serpiente mordiéndose la cola, y máquina de coser diseñada por Elias Howe (1819-1867) tras soñar con una tribu rodeándolo con lanzas.

A pesar de que la evidencia científica es todavía relativamente escasa, vimos que, lejos de ser un epifenómeno, los sueños tienen aparentemente la función de procesar, contextualizar y asimilar la información de lo que vivimos durante el día. La capacidad de manipular el sueño de los otros (o nuestros propios sueños) como en *Inception* es relativamente limitada y probablemente así sea mejor, pues si llegáramos a manipular los sueños quizás terminaríamos afectando la función clave que estos tienen en el funcionamiento de nuestro cerebro. Sin embargo, veremos que una predicción de la película es mucho más real y concreta. En vez de limitarse a hurgar los pensamientos de las personas —lo cual ya es mucho— Cobb y sus secuaces deciden probar algo aún más audaz: implantar una idea, una memoria. En el capítulo anterior argumentamos que la memoria define quiénes somos. ¿Será entonces posible implantar un recuerdo?

Y de ser posible, ¿qué consecuencias tendría para la identidad de la persona? Este es justamente el tema del próximo capítulo.

161 Esto recuerda a la famosa *paradoja de Aquiles y la tortuga* propuesta por Zenón de Elea: Aquiles corre contra una tortuga, pero le da una ventaja inicial porque esta es más lenta; cuando Aquiles llega hasta donde estaba la tortuga, esta ya se ha adelantado un poco, al llegar Aquiles al nuevo punto, la tortuga vuelve a estar un poco más adelante, y así sucesivamente, con lo cual Aquiles aparentemente nunca alcanza a la tortuga. Esta paradoja —como la idea de prolongar la vida de una persona sumergiéndola en un sueño más profundo en donde se retarda el tiempo— está basada en el hecho de que siempre hay infinitos puntos dentro de un intervalo de tiempo, por más pequeño que sea.

162 Por otro lado, debido a las violaciones del tiempo y las leyes de la física en los sueños, el trompo de Cobb queda girando eternamente cuando no está despierto.

163 Estos casos están descriptos en:

- A. Szucs, J. Janszky, A. Hollo, G. Migléczi and P. Halász. “Misleading hallucinations in unrecognized narcolepsy”. *Acta Psychiatrica Scandinava* 108: 314-317; 2003.
- E. Wamsley, C. Donjacour, T. Scammell, G. Lammers and R. Stickgold. “Delusional confusion of dreaming and reality in narcolepsy”. *Sleep* 37: 419-422; 2014.

164 Adaptado de: A. Rogers, M. Aldrich, C. Caruso. “Patterns of sleep and wakefulness in treated narcolepsy subjects”. *Sleep* 17:590–597; 1994.

165 E. Wamsley, C. Donjacour, T. Scammell, G. Lammers and R. Stickgold. Op. cit.

166 Para llegar a hacernos estas preguntas durante el sueño, debemos plantearnos las mismas preguntas mientras estamos despiertos. Por más ridículo que resulte, por ejemplo, verificar que no podemos atravesar la pared con la mano, al probar esto repetidamente durante la vigilia aumentamos las chances de eventualmente hacer lo mismo durante en el sueño.

167 William Moorcroft. *Understanding sleep and dreaming*. Springer, 2013 (capítulo 7.8).

168 El electrooculograma estima los movimientos oculares registrando el campo eléctrico que estos generan, usando electrodos colocados alrededor del ojo.

169 Figura adaptada de: Stephen La Berge. “Lucid dreaming: Psychophysiological studies of consciousness during REM sleep”. En R. Bootsen, J. Kihlstrom y D. Schacter (eds.), *Sleep and Cognition.* Washington, DC: APA Press (pp. 109-126; 1990).

170 Miguel Pais-Vieira, Mikhail Lebedev, Carolina Kunicki, Jing Wang and Miguel Nicolelis. “A brain-to-brain interface for real-time sharing of sensorimotor

information". *Scientific Reports* 3:1319; 2013.

171 Deirdre Barrett. "The 'committee of sleep': A study of dream incubation for problem solving". *Dreaming* 3: 115-122; 1993. Ver también: Robert Stickgold, April Malia, Denise Maguire, David Roddenberry and Margaret O'Connor. "Replaying the game: Hypnagogic images in normal and amnesics". *Science* 290: 350-353; 2000.

172 Los mecanismos de censura que desaparecieron en Phineas Gage tras su lesión en el lóbulo frontal, generando un radical cambio de personalidad (ver capítulo 1). En línea con esta evidencia, durante el sueño cae notablemente la activación del lóbulo frontal. Ver por ejemplo: R. Stickgold, J. Hobson, R. Fosse and M. Fosse. "Sleep, learning, and dreams: off-line memory reprocessing". *Science* 294: 1052-1057; 2001.

173 Para una discusión sobre este tema y potenciales fallas del psicoanálisis ver, por ejemplo: Giuliana Mazzoni and Elizabeth Loftus. "When Dreams Become Reality". *Consciousness and Cognition* 5: 442-462; 1996.

174 Eugene Aserinsky and Nathaniel Kleitman. "Regularly occurring periods of eye mobility and concomitant phenomena during sleep". *Science* 18: 273-274; 1953.

175 Más adelante, el período no-REM fue también subdividido en 4 fases de sueño, según patrones estereotipados en los registros electroencefalográficos: la fase 1, típica al adormecernos, es cuando tenemos un sueño ligero y podemos ser fácilmente despertados; la fase 2 suele darse como transición desde o hacia el sueño profundo de las fases 3 y 4 (cuando más nos cuesta despertarnos).

176 Matthew Walker; *Why we sleep*; Penguin; 2018.

177 William Dement. "The effect of dream deprivation". *Science* 131: 1705-1707; 1960.

178 Alan Hobson and Robert McCarley, R. "The brain as a dream state generator: an activation-synthesis hypothesis of the dream process". *American Journal of Psychiatry* 134: 1335–1348; 1977.

179 El insomnio prolongado no solo acarrea consecuencias catastróficas en la memoria, sino también en el estado general de salud de una persona. Se da en una enfermedad muy rara e incurable denominada insomnio familiar fatal. Los pacientes diagnosticados con esta enfermedad comienzan a tener un deterioro muy notable tras unas semanas de no poder dormir (semejando pacientes con demencia avanzada) y mueren a los pocos meses.

180 En una trágica ironía del destino, García Márquez padeció alzhéimer en los últimos años de su vida. Lamentablemente, aun no tenemos un brebaje mágico como el del gitano Melquíades para curar los efectos devastadores del verdadero "mal del olvido".

181 John Jenkins and Karl Dallenbach. "Obliviscence during sleep and waking". *American Journal of Psychology* 35: 605-612; 1924. Para una revisión de los efectos que tiene el dormir en la memoria ver:

- S. Diekelmann and J. Born. "The memory function of sleep". *Nature Reviews Neuroscience* 11: 114-126; 2010.

- R. Stickgold, J. Hobson, R. Fosse and M. Fosse. Op. cit.
- M. Walker and R. Stickgold. “Sleep-dependent learning and memory consolidation”. *Neuron* 44: 121-133; 2004.

182 O. Lahl, C. Wispel, B. Willigens and R. Pietrowsky. “An ultra short episode of sleep is sufficient to promote declarative memory performance”. *Journal of Sleep Research* 17: 3-10; 2008.

183 W. Plihal and J. Born. “Effects of early and late nocturnal sleep on declarative and procedural memory”. *Journal of Cognitive Neuroscience* 9: 534-547; 1997.

184 Para una revisión de estos resultados, ver: S. Diekelmann and J. Born. “The memory function of sleep”. *Nature Reviews Neuroscience* 11: 114-126; 2010.

185 R. Stickgold, J. Hobson, R. Fosse and M. Fosse. Op. cit.

186 Ibídem.

187 Para un desarrollo más elaborado de esta interpretación, ver:

- R. Stickgold, J. Hobson, R. Fosse and M. Fosse. Op. cit.
- M. Walker and R. Stickgold. “Overnight alchemy: sleep-dependent memory evolution”. *Nature Reviews Neuroscience* 11: 218-219; 2010.
- Matthew Walker; op. cit. En su fascinante libro, Walker argumenta que otra función del sueño es la de procesar emociones que vivimos durante el día, algo crucial para desarrollar una conducta social (ver capítulo 10).

IX. Total Recall

Manipulación e implante de memoria

Douglas Quaid es un obrero de la construcción que sueña con ir a Marte. Buscando escapar de la monótona rutina de su vida, y tentado por un anuncio en el subte, decide visitar las oficinas de Rekall para realizar un viaje virtual al planeta rojo mediante un implante de memoria. El empleado de Rekall le nombra una variedad de ventajas que tiene realizar el viaje virtual comparado con ir físicamente a Marte: es más barato y seguro, con vuelos en primera clase, los mejores hoteles y sin las complicaciones de unas vacaciones en la vida real. Quaid duda y se pregunta qué tan creíbles serán las experiencias del viaje; pero el empleado le asegura que las vacaciones serán tan reales como cualquier memoria en su cabeza, ya que su cerebro no notará la diferencia.

Arreglando los detalles de su viaje, Quaid rechaza automáticamente las opciones adicionales ofrecidas, hasta que el empleado le pregunta por aquello que se mantiene exactamente igual en cada una de sus vacaciones. Quaid no encuentra respuesta y finalmente el vendedor le revela que lo que siempre se repite es él mismo, pues no deja de ser la misma persona en cualquier lugar o circunstancia. La opción que en definitiva le resultará irresistible es el “viaje del Ego”, en donde cambia su personalidad. A fin de cuentas, todo es un implante de memoria, y nada impide que Quaid pueda ser un magnate, una leyenda del deporte o un agente secreto, como finalmente elige ser.

Todo está listo para que Quaid disfrute su viaje a Marte, pero una vez iniciado el procedimiento algo anda mal. Comienza a recuperar recuerdos perdidos y eventualmente termina descubriendo que su nombre no es Quaid, sino Hauser, un agente secreto en Marte, y que tanto su identidad como su pasado y todos los recuerdos de estar varios años casado con su mujer no son más que un implante de memoria. Finalmente logra viajar a Marte para terminar involucrado en una monumental trama enfrentando a una corporación que

especula con racionalizar y cobrar el uso del aire.

Después de varias persecuciones y tiroteos, un doctor se presenta en su hotel y le dice que en realidad no está en Marte, sino en una cama en Rekall, viviendo un implante de memoria que se les está yendo de las manos, y que él mismo fue introducido artificialmente en su mente para rescatarlo de lo que empieza a ser una pesadilla. Pareciera tener sentido, pues su vida dio un vuelco dramático justo después de haber comenzado sus vacaciones virtuales. Aunque, por otro lado, también tiene sentido que el implante de las vacaciones en Marte despertara memorias parcialmente borradas y que sus deseos de conocer Marte no sean más que recuerdos aflorando de su subconsciente.

Este es, a grandes rasgos, el dilema que plantea *Total Recall*[188]: ¿quién es la persona que enfrenta todas estas situaciones? Será tal vez Quaid, un simple obrero viviendo la realidad ilusoria de ser un agente secreto, o Hauser, un agente secreto a quien le implantaron artificialmente la identidad de Quaid para borrar su pasado. La ambigüedad de no poder elegir entre Quaid y Hauser domina el resto de la trama y las posibles interpretaciones de la película.

Una vez que es atrapado, el dueño de la corporación a la que enfrenta le muestra un video de sí mismo en el pasado con la personalidad de Hauser, quien le confirma que su memoria actual es un implante al que está a punto de someterse en el momento de filmar el video. Hauser también alega que llegado el momento querrá restablecer la memoria original y volver a ser quien era. Quaid se resiste, pues en el fondo desprecia a Hauser, es decir, desprecia a alguien que tiene su mismo cuerpo pero que no es él porque tiene otra memoria.

Quisiera una vez más referir a Borges. En “La memoria de Shakespeare” relata las desventuras de Hermann Soergel, un profesor de literatura en la época victoriana, a quien un colega, tras una larga noche en una taberna inglesa, le ofrece ni más ni menos que la memoria de Shakespeare. Soergel acepta algo incrédulo, aunque anhelante de poder compartir la mente del escritor más brillante de todos los tiempos. Poco a poco va accediendo a los pensamientos privados del bardo, mas tras la fascinación inicial termina rechazando

la memoria de Shakespeare, al notar que comienza a confundirse con la suya propia y que paulatinamente va perdiendo su identidad. Dice Soergel en el cuento de Borges:

> En la primera etapa de la aventura sentí la dicha de ser Shakespeare; en la postrera, la opresión y el terror. Al principio las dos memorias no mezclaban sus aguas. Con el tiempo, el gran río de Shakespeare amenazó, y casi anegó, mi modesto caudal. Advertí con temor que estaba olvidando la lengua de mis padres. Ya que la identidad personal se basa en la memoria, temí por mi razón.

Como viéramos en el capítulo 7, en el siglo XVII John Locke se preguntó por aquello que determina la identidad a partir de la *parábola del príncipe y el zapatero*. Locke, como Borges (o la trama de *Total Recall*), afirma que la identidad va ligada a la memoria. Los recuerdos constituyen a una persona. Por mi parte, soy aquel que estudió en la Universidad de Buenos Aires, que luego hizo un doctorado en Lübeck y que ahora escribe este libro en Inglaterra. ¿Pero qué tan fehacientes son los recuerdos que me llevan a ser quién soy? ¿Duró realmente ocho horas aquel examen de Mecánica Clásica en el que nos llevábamos una vianda para no desfallecer mientras seguíamos tratando de resolver problemas interminables? ¿Surgió de aquella charla en el Colegio Nacional de Buenos Aires la idea de escribir este libro?

Si hay algo de lo que estoy seguro es de mi propia existencia, de mi pasado y de saber quién soy. Pero antes mencionaba lo falaz de uno de los recuerdos de mi infancia, el de la pelota naranja en el partido de Argentina contra la Unión Soviética (ver capítulo 4). Me pregunto entonces qué tan frágil y maleable es mi memoria y, consecuentemente, mi identidad. Me pregunto, asimismo, si alguien podría implantarme recuerdos y cambiar mi persona como en *Total Recall*.

En la mitología griega, la memoria era personificada en la diosa Mnemósine, madre de las nueve musas[189]. ¿Pero por qué tener una diosa de la memoria? Es entendible que haya una diosa de la caza (Artemisa), de la sabiduría (Atenea), del amor y la belleza (Afrodita) o

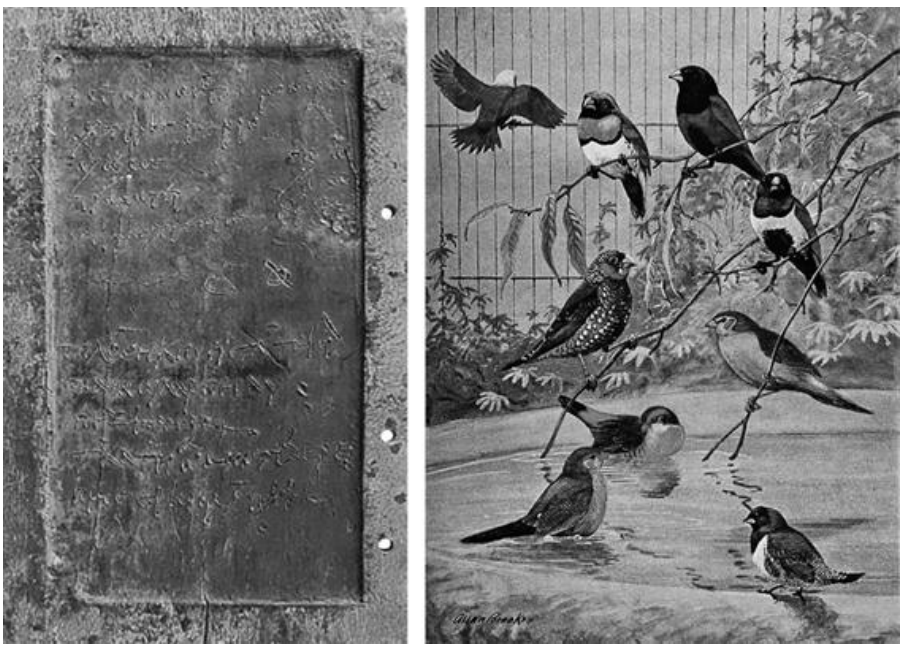
de la Tierra (Gaia). Pero ¿de la memoria? ¿Por qué no de las sensaciones, de los impulsos o la conciencia?

Uno de los diálogos de Platón relata cómo Critias invoca a Mnemósine antes de una larga exposición porque teme no recordar lo que tiene que decir (Platón, *Critias*). Y no es de despreciar la importancia de la memoria en la Antigüedad, pues, a falta de papel, los oradores dependían exclusivamente de su memoria para poder dar un discurso; tal es así, que dos de los tres libros fundamentales sobre la memoria de la Antigüedad refieren a la oratoria: *De oratore*, por Cicerón, e *Institutio oratoria*, por Quintiliano (el tercero es *Rhetorica ad Herennium,* de autor anónimo).

Mnemósine (óleo de Dante Gabriel Rossetti) y tres de las nueve musas (óleo de Eustache Le Sueur): Clío (historia), Euterpe (música) y Thalía (comedia).

La manufactura del papel se origina durante la dinastía Han, a principios del siglo II en China, pero tardaría más de un milenio en llegar a Occidente. Se expande al mundo musulmán a partir de la batalla de Talas (751 d. C.) entre el califato árabe y el imperio chino, en la cual los árabes toman prisioneros que les enseñan a procesar

papel a cambio de su libertad. Recién en el siglo XII llegaría a Europa la técnica de la fabricación del papel, a partir la reconquista del sur de España. Hasta ese entonces, las escrituras se hacían en papiros —muy costosos por ser importados de Egipto— o en los aún más onerosos pergaminos, hechos con cuero de animales. En la Antigüedad, particularmente en la antigua Grecia, también se usaban las inscripciones en tablas de cera (tablas de madera con la superficie encerada); pero estas técnicas de escritura eran muy engorrosas y caras, y de allí su uso limitado.



Representación de la memoria según Platón: como el grabado en una tabla de cera (izquierda) o una jaula de pájaros (derecha)[190].

Platón describe a la memoria como la marca impresa en un bloque de cera, alegando que cuando la cera en la mente de una persona es pura, profunda y abundante, las memorias tienen suficiente lugar para establecerse y ser duraderas (*Teeteto* 34). En el mismo tratado considera a la memoria como una jaula de pájaros. La jaula está vacía en un niño y a medida que este crece se va llenando de pájaros, representando cada memoria o conocimiento. De esta manera, Platón diferencia entre poseer conocimiento y acceder a este (lo que en

nuestros días llamamos *codificación* y *recuperación* de la memoria). Poseer conocimiento es tener la jaula llena de pájaros, mientras que acceder al conocimiento implica poder tomar cualquiera de estos pájaros, examinarlo y soltarlo en la jaula de nuevo. Estas dos imágenes de Platón reflejan la idea intuitiva que tenemos sobre la memoria, aunque distan mucho de la realidad. Lejos de ser grabados fijos en la cera o pájaros en una jaula que aguardan ser usados, la formación y almacenamiento de los recuerdos es un proceso mucho más dinámico.

El estudio experimental y sistemático de la memoria comienza con el psicólogo alemán Hermann Ebbinghaus, quien a fines del siglo XIX llevó a cabo una serie de simples, aunque tediosos experimentos consigo mismo. Construyó 2.300 palabras sin sentido, cada una de ellas formada por tres letras (dos consonantes con una vocal en el medio) y en infinidad de experimentos evaluó cuántas de estas palabras tomadas al azar recordaba tras distintos intervalos de tiempo[191].

Ebbinghaus determinó dos principios fundamentales. Primero, mientras algunos recuerdos duran no más de unos pocos minutos, otros persisten horas, semanas o incluso años; esto es lo que actualmente se conoce como memoria de corto y de largo plazo. La *memoria de corto plazo* mantiene información por breves intervalos de tiempo y me lleva a ser consciente de la sucesión de acontecimientos de mi presente. Es la que uso, por ejemplo, para acordarme de lo que quiero decir en este momento mientras elijo las palabras adecuadas. La *memoria de largo plazo*, en cambio, está constituida por cosas que recuerdo de mi pasado, como la memoria de un buen vino, la cara de mi madre o un día en la playa.

Casi toda la memoria de corto plazo termina perdiéndose en el olvido y solo una ínfima parte de lo que me acontece en este momento pasará a constituir mi memoria de largo plazo. Pero ¿qué es lo que hace que algunas memorias sean efímeras y que otras duren toda una vida? Este es el segundo principio de Ebbinghaus: la repetición y la práctica hacen perdurar los recuerdos, pues al evaluar sus resultados notó que cuanto más repetía una lista de palabras, por más tiempo la

recordaba. La memoria de largo plazo está formada justamente por hechos salientes a los que prestamos atención, recuerdos que traeremos infinidad de veces a nuestro consciente y que vamos consolidando con cada repetición.

Usando experimentos similares a los de Ebbinghaus, los psicólogos alemanes Georg Müller y Alfons Pilzecker observaron a fines del siglo XIX que el aprendizaje de una lista de palabras era interrumpido si pedían a los sujetos que aprendieran otra lista inmediatamente después de la primera, aunque no tanto si aprendían la segunda lista un tiempo más tarde. A partir de estos resultados concluyeron que hay un período crítico en el cual el aprendizaje de la segunda lista interfiere con la consolidación de la primera. Este período de consolidación explica la amnesia temporaria que tienen algunos boxeadores tras un nocaut, haciendo que a veces ni siquiera recuerden haber ido a la pelea. La amnesia se da por la pérdida de consciencia debido al golpe recibido, la cual interrumpe la consolidación de los recuerdos que la precedieron. Sin embargo, más allá de la existencia de un período crítico para la fijación de la memoria, la consolidación es un proceso constante. Por ejemplo, solemos tener recuerdos más vívidos de aquellos eventos de los cuales tenemos fotografías que miramos de vez en cuando. Las fotos nos llevan a repasar los hechos y a consolidar una y otra vez los recuerdos relacionados con ellos.

El modelo de consolidación ha sido una piedra angular en nuestro conocimiento sobre el funcionamiento de la memoria durante casi 100 años: cuando recién es adquirida, la memoria es lábil y está representada por conexiones específicas entre neuronas que, si llegan a consolidarse (por ejemplo a través de la repetición), harán que la memoria perdure en el tiempo[192]. Siguiendo la analogía de Platón, si los recuerdos son como grabados en la cera, cuanto más evocamos un recuerdo más profundo será el grabado.

Modelo clásico de consolidación de la memoria. Si vemos por primera vez a Luke Skywalker (en *La guerra de las galaxias*) inicialmente formamos un recuerdo lábil de este, representado por un grupo de neuronas interconectadas. Al repetir y consolidar la memoria de Luke, las conexiones se vuelven más fuertes y afianzan la memoria.

Parecería, entonces, que no hay demasiado margen para la posibilidad de borrar e implantar memorias como en *Total Recall*, pues, según la teoría de consolidación, el único momento en el que podría manipularse una memoria es justo después de adquirida, porque una vez consolidada es perenne. Sin embargo, en 1968 un grupo de científicos norteamericanos publicó una serie resultados que no encajaban con esta visión[193].

Seguramente hemos escuchado hablar de los reflejos condicionados de Pavlov: si a un perro se le da de comer justo después de hacer sonar una campanilla, el perro asocia la campanilla a la comida y basta con hacerla sonar para que empiece a salivar esperando su alimento. Una variación de este experimento es el *condicionamiento por miedo* (*fear conditioning*), muy usado en ratas porque lo aprenden casi inmediatamente. En este caso, se presenta un sonido específico seguido por un pequeño shock de electricidad que incomoda al animal y, tras unas pocas repeticiones, es notorio que tras escuchar el sonido la rata se prepara para recibir la descarga. Por otro lado, seguramente también habremos escuchado acerca del electroshock, que consiste en aplicar una fuerte descarga eléctrica para tratar ciertas patologías y que tiene un efecto similar a la trompada de nocaut, generando una

amnesia temporaria.

Juntando estas dos piezas de información, el experimento que demolió la teoría clásica de consolidación fue muy sencillo. Primero los investigadores crearon un reflejo de miedo condicionado en ratas: la presentación de un ruido era seguida por una pequeña, aunque molesta, descarga eléctrica. Los animales aprendieron rápidamente a asociar ambos estímulos y adaptaron su comportamiento, tomando mucho menos agua de lo normal debido a su estado de ansiedad. Inmediatamente después de haber creado el reflejo condicionado, se le aplicó a un grupo de estas ratas una gran descarga eléctrica (electroshock) para interferir la consolidación de esta memoria, y cuando repitieron el experimento al día siguiente, notaron que no recordaban la sesión de aprendizaje del día anterior y tomaban agua normalmente. Pero lo más interesante fue lo que hicieron con un tercer grupo de ratas, a las que les aplicaron el electroshock no tras al aprendizaje en el primer día, sino tras volver a hacer la misma tarea un día después; es decir, y he aquí la clave, una vez que aprendieron y consolidaron el reflejo condicionado. El resultado sorprendente fue que esto hizo que olvidaran lo que habían aprendido el día anterior y volvieran a tomar agua normalmente. En otras palabras, al reactivar e interferir una memoria ya consolidada, lograron borrarla.

Tan predominante era la teoría de consolidación en el siglo pasado, que estos resultados pasaron mayormente desapercibidos hasta que con un experimento similar, usando manipulaciones farmacológicas con ratas, se llegó a la misma conclusión 30 años más tarde[194]. En la actualidad sabemos entonces que la consolidación es un proceso dinámico: no existen los grabados estáticos en la cera; los recuerdos van cambiando en un constante proceso de *reconsolidación*[195]. Por lo tanto, los recuerdos son lábiles no solo cuando son adquiridos (como afirma la teoría de la consolidación) sino también cada vez que son evocados. El hecho de traer una memoria al consciente implica cambiarla.

La noción de reconsolidación, de poder modificar o incluso borrar ciertas memorias, tiene connotaciones terapéuticas extraordinarias. Muchas personas sufren de *trastorno de estrés postraumático* causado

por una experiencia dramática como una violación, un accidente automovilístico o situaciones extremas vividas en la guerra. Quienes sufren de este trastorno no pueden olvidar esas experiencias y su evocación, involuntaria e indeseada, genera estrés y pánico. El gran problema es que estos pacientes recién acuden al profesional cuando las reacciones ya se han consolidado en el cerebro, e idealmente uno querría intervenir en el proceso de consolidación de memorias indeseadas ni bien acontecen, algo que es obviamente imposible de implementar en la práctica. Sin embargo, si la memoria entra en un estado lábil cada vez que es traída al consciente, podemos entonces evocar el recuerdo indeseado y tratar de borrarlo.

Implementar los experimentos anteriores en humanos no es tan fácil, pues por un lado el electroshock solo es usado en casos muy excepcionales, dado que presenta contraindicaciones importantes y, por el otro, las drogas que fueron usadas en los estudios con animales son altamente tóxicas, aunque tal vez drogas menos nocivas tengan el mismo efecto.

El propranolol es un remedio para la presión tomado por millones de pacientes todos los días. Es lo que se conoce como un bloqueador beta, cuya función principal es disminuir la frecuencia cardíaca. Sin embargo, el propranolol también tiene el efecto secundario de bloquear la síntesis de proteínas involucradas en mantener las conexiones entre neuronas y, como ilustráramos con Luke Skywalker, la conectividad entre las neuronas es la base de la memoria. Merel Kindt, una investigadora holandesa, probó entonces usar propranolol para interrumpir la reconsolidación de memorias en humanos[196]. El experimento se basaba en establecer un reflejo condicionado (la presentación de una foto era seguida por una incómoda descarga eléctrica), luego evocar el reflejo al día siguiente tras darle propranolol a los sujetos, y finalmente evaluar la reacción a las fotos en el tercer día. Kindt y sus colaboradores observaron que la aplicación de propranolol efectivamente lograba borrar el reflejo condicionado.

A pesar de lo promisorio de estos resultados, hay un mundo de diferencia entre eliminar la reacción a una pequeña descarga eléctrica

y borrar una memoria traumática del pasado. Mientras algunos trabajos afirman que es posible paliar trastornos de estrés postraumático al hacer que los pacientes revivan esas experiencias tomando propranolol, otros tantos concluyen que la utilidad clínica de este procedimiento es relativamente limitada[197]. Cabe aclarar que, aun en el mejor de los casos, la reconsolidación de recuerdos traumáticos no borra la memoria en sí misma, sino la reacción emocional del sujeto. Es decir, un veterano de guerra seguirá recordando por el resto de sus días situaciones dramáticas que haya vivido, pero la idea es que pueda hacerlo sin sobreestresarse o entrar en pánico. Eliminar completamente una memoria es algo mucho más controversial desde el punto de vista ético, pues nuestros recuerdos, por más dramáticos que sean, constituyen nuestra vida y personalidad; el aprendizaje al que recurrimos para tratar de no repetir malas experiencias.

En general es casi imposible eliminar totalmente recuerdos tan afianzados como los de una memoria traumática (reiteramos que los trabajos anteriores solo eliminan la reacción emocional que estas generan), pero probablemente podamos alterar memorias más vagas y lábiles, y sin uso de droga alguna, solo a través de la sugestión. La gran experta en el tema es la psicóloga Elizabeth Loftus, quien en un famoso experimento demostró cómo la reconsolidación de la memoria puede llegar a la fabulación o a la formación de falsos recuerdos[198].

Loftus mostró a distintos sujetos el video de un accidente de tránsito, para luego pedirles que estimaran la velocidad de los autos. Lo interesante fue que mientras a un grupo de sujetos le preguntó por la velocidad de los autos cuando *chocaron,* a otro grupo le hizo la misma pregunta, pero usando la palabra *colisionaron;* con un tercer grupo utilizó la palabra *golpearon;* con el cuarto grupo, *contactaron,* y al último grupo le preguntó por la velocidad de los autos cuando *se estrellaron.* Todos los sujetos vieron exactamente el mismo video y en las mismas condiciones, pero aquellos con los cuales usó la palabra *estrellaron,* dieron la estimación de velocidad más alta, seguidos por aquellos con los que usó *colisionaron,* luego *golpearon*, *chocaron* y finalmente, *contactaron*. Aún más sorprendente fue el hecho de que

una semana después les preguntó a las mismas personas si recordaban haber visto vidrios rotos en la escena del accidente y mientras un 32% de los sujetos con los que usó la palabra *estrellaron* respondieron (erróneamente) que sí, solo un 14% con los que usó la palabra *chocaron* afirmaron haberlos visto.

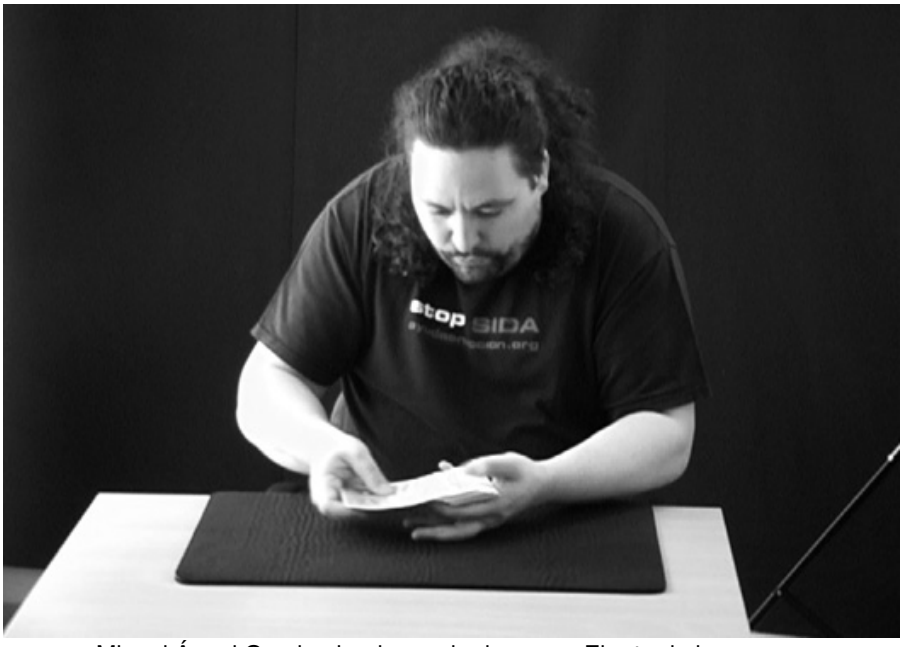
Los resultados de Loftus ponen de manifiesto lo frágil que es nuestra memoria y cómo puede ser manipulada en el proceso de reconsolidación, con sólo cambiar una palabra en una pregunta[199]. Pero más allá de alterar la impresión de algo que hayamos visto o vivido, ¿podremos hacer uso de la sugestión para llegar a implantar el recuerdo de algo que nunca sucedió?

Siguiendo con esta línea de experimentos, Loftus dio a un grupo de personas descripciones de cuatro eventos de su infancia provistos por sus familiares. Pero lo que no sabían es que uno de estos recuerdos —el haber estado perdidos en un centro comercial— era completamente falso. Loftus pidió a los sujetos que escribieran lo que recordaban sobre estas historias durante cinco días consecutivos, y aunque en un principio dudaban del evento en el centro comercial, terminaron convencidos de su veracidad. Tras cada recapitulación, uno de estos sujetos, por ejemplo, alegaba recordar cada vez más detalles de aquel día: lo amable que era el señor con una camisa azul, gafas y algo pelado que lo ayudó; el miedo que sintió de no volver a ver a su familia, y a su madre regañándolo al encontrarlo; todos recuerdos falsos. Finalmente le revelaron que una de las cuatro memorias era falsa y al pedirle que adivinara cuál de estas era, eligió una memoria real, sin poder creer que nunca había estado perdido en un centro comercial. Asimismo, Loftus relata cómo psicoanalistas desvergonzados hicieron uso de la sugestión y la reconsolidación para llevar a pacientes a creer que habían sufrido abusos sexuales y otros eventos dramáticos en su infancia (estos pacientes luego ganarían juicios millonarios al demostrarse que estos recuerdos no eran más que un engaño implantado en su memoria)[200].

Vimos anteriormente que la memoria es un proceso creativo, que recordamos muy poca información y que reconstruimos nuestros recuerdos usando inferencias, generando los esquemas de Bartlett, la

base de nuestros recuerdos. Ahora también vemos que esta construcción es un proceso muy dinámico, que va cambiando con el tiempo, lo cual permite manipular e incluso implantar recuerdos.

Más allá del indiscutido mérito de científicos como Elizabeth Loftus, a mi entender los grandes conjuradores, los grandes manipuladores de recuerdos, son los magos, quienes con su arte nos llevan a experimentar lo imposible. Por ejemplo, los magos pueden recapitular lo que ha pasado antes de revelar un truco, pero esta recapitulación suele ser parcial porque en ella refuerzan algunos hechos que quieren sean recordados, e ignoran otros tantos que prefieren sean olvidados, ya que pueden dar indicios de cómo se llevó a cabo el truco. Pero el mago avezado no solo influenciará sino que también alterará el contenido de esos recuerdos al ser reconsolidados. Podrá por ejemplo mezclar y manipular rápidamente las cartas y luego dárselas a un espectador para que haga varios cortes (que no tienen consecuencia alguna). Tras varios pases y distracciones, podrá sintetizar lo acontecido diciendo algo así como “las cartas han sido mezcladas”, mirando al espectador mientras sostiene el mazo. Esta es la semilla de la falsa memoria; una afirmación vaga pero correcta, ya que no especifica quién hizo la mezcla, aunque sugiere la idea de que ha sido el espectador. Si en lugar de una sugerencia indefinida, el mago usara una afirmación más precisa como “el espectador ha mezclado las cartas”, entonces la audiencia podría notar inmediatamente el engaño. El talento del mago está en reforzar una y otra vez estas ideas (pero sin decirlo explícitamente), para terminar implantando el recuerdo que da lugar a algo imposible, por ejemplo, que pueda encontrar una carta en un mazo que no ha manipulado[201].



Miguel Ángel Gea haciendo magia de cerca. El arte de los magos no solo se basa en manipular las cartas sino también la atención de los espectadores y el recuerdo de lo que hacen en sus trucos.

Hace medio siglo, en los lejanos Urales vivía Akakhi Akakhievitch, un brillante, aunque ignoto, neurocirujano ruso. Convencido de que las ideas están representadas por neuronas específicas, comenzó a hacer experimentos en animales (y en trotskistas esperando su sentencia de muerte en Siberia) tratando de localizar aquellas que codificaban el concepto de "madre". Tras varios intentos encontró neuronas respondiendo únicamente a la madre de los sujetos, sin importar como la mostrara: de frente o de perfil, boca abajo, en diagonal, en una caricatura, etcétera. Quiso el destino que en esos días Akakhievitch recibiera un paciente que le imploraba poder olvidar a su autoritaria e insufrible madre. Haciendo uso de sus recientes conocimientos, el cirujano procedió a extirpar todas y cada una de las miles de neuronas que representaban el concepto "madre" en el cerebro del paciente. Cuando este despertó de la anestesia, había perdido efectivamente la noción de su madre; todos los recuerdos relacionados con ella se habían esfumado. Lleno de júbilo por el éxito de la operación, Akakhievitch decidió encarar el siguiente reto: encontrar las "neuronas de la abuela".

La historia, por supuesto ficticia, fue pergeñada por el neurocientifico Jerry Lettvin para amenizar un curso que daba a estudiantes del Massachusetts Institute of Technology a fines de la década del 60[202]. Inicialmente tomada como una broma, la crónica de Akakhievitch terminó desencadenando un intenso debate en neurociencia. Aunque la discusión sobre la existencia de neuronas representando cosas específicas es mucho más antigua y se remonta a William James, padre de la psicología moderna, quien a fines del siglo XIX postulara que "entre todas las células puede haber una central o pontificia a la que está ligada nuestra consciencia"[203].

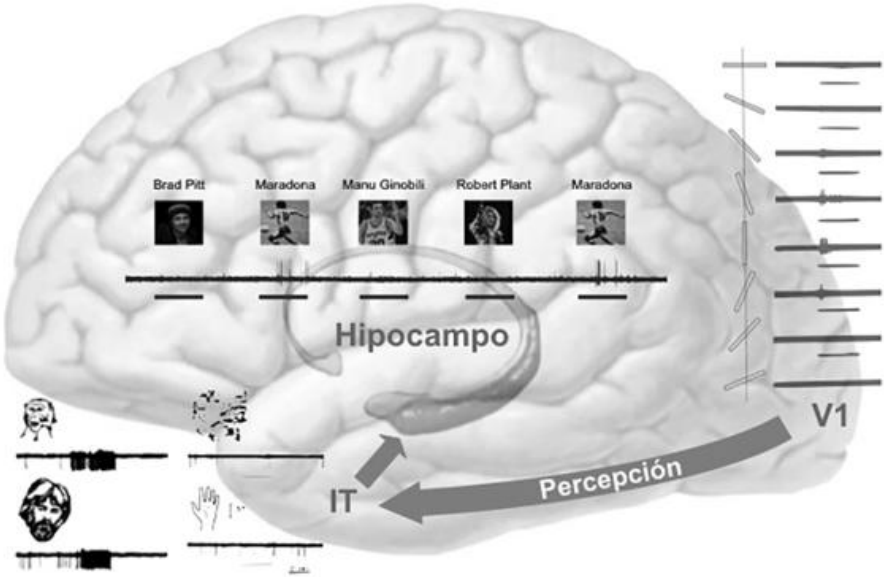
La idea de *células pontificias* pasaría desapercibida por décadas, pero sería luego retomada por el premio Nobel Charles Sherrington, quien argumentara que para decidir si, por ejemplo, estamos viendo la cara de un amigo, nuestra madre o una manzana, el cerebro no usa un solo pontífice —es decir, neuronas únicas— sino una *democracia de millones*[204]. En otras palabras, según Sherrington la representación de los conceptos no viene dada por la activación de neuronas específicas, sino que está distribuida en grandes poblaciones de neuronas. Estas neuronas responden a distintos detalles menores, y al sumar todos estos detalles (por ejemplo, la nariz, color de piel, ojos, pelo, labios, etcétera) se genera la representación del concepto.

Para fundamentar con evidencia científica la posibilidad de que existan o no "neuronas de la abuela" necesitamos hablar un poco sobre la percepción visual. La información captada por los ojos se transmite a la corteza visual primaria, en donde David Hubel y Torsten Wiesel descubrieron, en la década del 50, neuronas que responden a líneas con distintas orientaciones, resultados que los harían merecedores del Premio Nobel en 1981. Esta información llega al lóbulo temporal inferior (IT), sitio en el que, haciendo experimentos con monos, el investigador norteamericano Charles Gross encontró neuronas que responden específicamente a rostros. (En esta misma área, Doris Tsao y Winrich Freiwald luego descubrirían los *face patches* de los que hablamos en el capítulo 5).

A medida que avanzamos en el procesamiento de información visual en distintas áreas del cerebro, la información representada por cada

neurona se vuelve más compleja. Considerando este procesamiento jerárquico, a fines de la década del 60 el psicólogo polaco Jerzy Konorski se preguntó si tal configuración no terminaría en neuronas individuales representando objetos o personas (como postulara James décadas atrás y como un par de años después ridiculizara Lettvin con la historia de Akakhi Akakhievitch)[205]. De hecho, fue tras una estancia en el laboratorio de Konorski que Charles Gross tuvo la idea de investigar respuestas a caras en áreas visuales superiores. Pero las neuronas descubiertas por Gross distan mucho de ser aquellas que predijera Konorski, pues responden a caras en general y no a personas específicas.

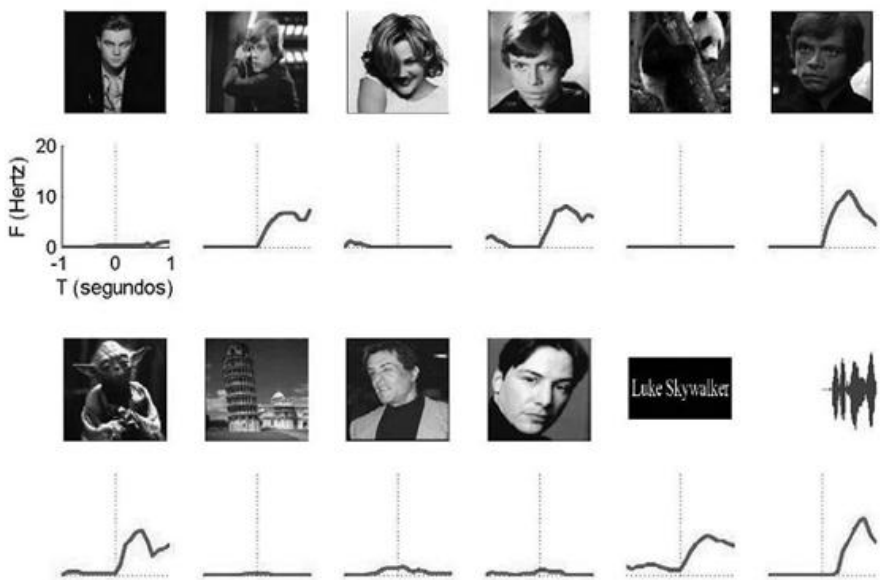
Las neuronas de las áreas visuales superiores proyectan al hipocampo, en donde en capítulos anteriores mencionaba que —haciendo experimentos en pacientes con epilepsia, implantados con electrodos intracraniales por motivos clínicos— encontramos neuronas que responden a personas específicas, como Jennifer Aniston, Halle Berry o Diego Maradona. Es decir, estas neuronas no responden a una foto en particular de una persona, sino al concepto. Por ejemplo, la neurona de Luke Skywalker respondió a tres fotos distintas de él e incluso a su nombre escrito y dicho por una voz sintetizada en la computadora. Nada más cercano a las “neuronas de la abuela” de Lettvin —células que responden solo a una persona, sin importar como se vea—, pero debemos considerar más cuidadosamente esta afirmación.



Procesamiento de información visual. Las neuronas en la corteza visual primaria (V1, en la parte posterior del cerebro) responden a líneas con una determinada orientación (en este caso, una línea vertical). Esta información se transmite a áreas visuales superiores (a través de la llamada ruta de percepción), para terminar en la corteza inferotemporal (IT), en donde se encontraron neuronas que responden a caras y no a otros estímulos visuales. La información de IT se transmite directamente al hipocampo en donde hay neuronas que responden a conceptos específicos; en este caso, Maradona[206].

Me detengo en los detalles de esta discusión porque están muy ligados a la posibilidad de borrar o manipular artificialmente recuerdos y, más en general, al funcionamiento de la memoria. Ante todo, si alguna vez encontramos una neurona respondiendo a Jennifer Aniston, esta es una de tantas, porque la posibilidad de dar con una única neurona (entre billones) respondiendo a un concepto específico es prácticamente cero. O sea, si encontramos una neurona respondiendo a Aniston, tiene que haber más, formando lo que se conoce como un *cell assembly* o *engrama*, un grupo de neuronas interconectadas que representan una memoria. Por otro lado, aun aceptando que hay tal vez miles de neuronas respondiendo a Jennifer Aniston, es imposible probar que estas neuronas responden solo a ella

y a ninguna otra cosa. Para demostrar esto, tendríamos que haber mostrado fotos de cada una de las personas que conocían los pacientes con los cuales hicimos estos registros, lo cual es imposible de hacer en la práctica. Más aún, el día después de haber encontrado la neurona de Jennifer Aniston, repetimos el experimento incluyendo fotos de los otros actores de *Friends*, la serie de televisión que la hizo famosa. La neurona volvió a responder a Aniston, como en el día anterior, pero también respondió a Lisa Kudrow, quien actuara en la misma serie. Asimismo, la neurona que se activaba con Luke Skywalker también respondió a una foto de Yoda —otro personaje de *La guerra de las galaxias*—, y a lo largo de los años hemos observado infinidad de ejemplos en donde estas neuronas responden a más de un concepto; aunque si este es el caso, los conceptos a los que responden suelen estar relacionados[207].



Respuestas de una neurona a tres fotos de Luke Skywalker y a su nombre escrito y dicho por la computadora (paneles en el extremo inferior derecho). La neurona también respondió a Yoda, otro personaje de *La guerra de las galaxias.*

Ya vimos que las neuronas de concepto se encuentran en el hipocampo, un área clave para la memoria, y que tiene sentido que así

sea porque tendemos a recordar conceptos —el significado de las cosas— y a olvidar detalles. Recuerdo, por ejemplo, un día con mi hermano en una playa en Brasil, pero no recuerdo cómo se veía en ese día, el color de su remera o el de su traje de baño. Todos esos detalles se perdieron en el olvido. En general tendemos a recordar conceptos: hermano, playa, Brasil. Mas ahora agregamos otra pieza de información al rompecabezas: estas neuronas no solo codifican conceptos sino también asociaciones entre ellos, ya que suelen responder a conceptos relacionados. Pero ¿por qué? ¿Por qué una neurona que se activa con distintas fotos de Luke Skywalker también responde a Yoda? ¿Por qué la neurona de Jennifer Aniston también respondió a Lisa Kudrow?

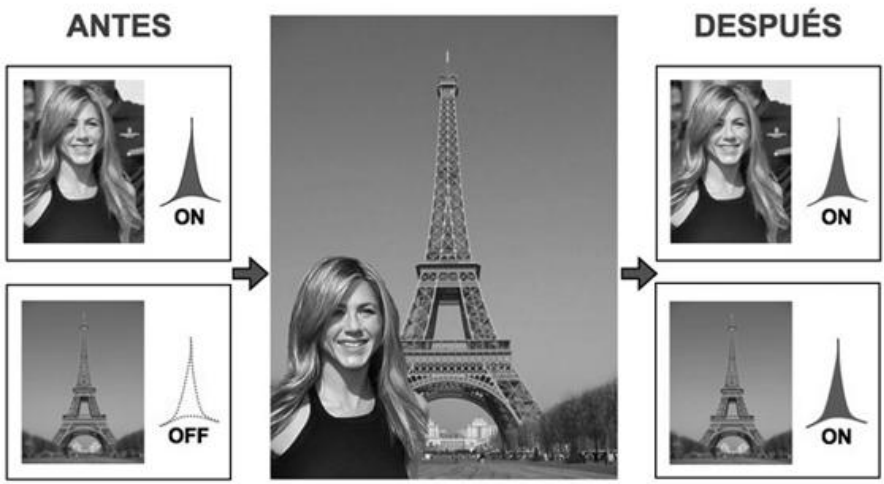
William James y mucho antes Aristóteles, en *De memoria et reminiscentia*, argumentaron que las asociaciones constituyen un mecanismo fundamental de la memoria. Pensar es abstraer, asociar conceptos, crear relaciones y categorías. Estas asociaciones son las que justamente nos ayudan a afianzar y evocar recuerdos. Al generar asociaciones y contexto tal vez no recordemos un hecho específico, pero muy probablemente recordaremos algún hecho asociado y a partir de allí llegaremos al que estábamos buscando. Sin embargo, las asociaciones son mucho más que un ayuda memoria; son la base de nuestros recuerdos.

A partir de la codificación de asociaciones, quisiera entonces proponer un modelo de cómo guardamos los recuerdos de nuestras experiencias en el hipocampo[208]. Supongamos que tenemos un grupo de neuronas que codifican el concepto de Luke Skywalker y otro que codifica el concepto de Yoda. Estos personajes están obviamente relacionados por ser parte de la misma película y tal asociación se representa teniendo algunas neuronas respondiendo a ambos conceptos.

Dos grupos de neuronas representando los conceptos de Luke Skywalker y Yoda. La asociación entre estos dos personajes, al ser parte de la misma película, está dada por neuronas que responden a ambos conceptos.

Técnicamente (al menos por ahora) no podemos, como en *Total Recall*, manipular la actividad de las neuronas en el hipocampo humano y ver cómo esto afectaría la memoria de la persona. Sin embargo, podemos hacer lo contrario: podemos manipular la memoria y ver si tal manipulación afecta las respuestas de las neuronas. Más concretamente, hace unos años hicimos un experimento en el que creamos artificialmente asociaciones mientras registrábamos la actividad de estas neuronas. Supongamos, por ejemplo, que tenemos una neurona que responde a Jennifer Aniston y no a la Torre Eiffel, y que construimos una figura como la de abajo, en la que se combinan los dos conceptos; una especie de *selfie* generando la memoria (ficticia) de ver a Jennifer Aniston en la Torre Eiffel.

A partir del momento en que los sujetos aprendieron esta asociación (entre otras tantas que debían recordar), vimos que la neurona que inicialmente respondía solo a Jennifer Aniston comenzó también a responder a la Torre Eiffel (sin Aniston en la foto)[209]. Lo más interesante fue la velocidad con que estas neuronas cambiaron sus respuestas, a veces tras mostrar una sola vez la "*selfie*", lo cual refuerza la idea de cómo estas neuronas están involucradas en la formación de recuerdos: de hecho, no vivimos nuestras experiencias una y otra vez para poder recordarlas y podemos recordar un evento saliente, que acontece una única vez y en tan solo segundos, por el resto de nuestras vidas. Si bien es cierto que los recuerdos son continuamente consolidados —o mejor dicho reconsolidados— hacen falta neuronas, como las neuronas de concepto en el hipocampo, que rápidamente puedan codificar estas asociaciones ni bien acontecen.



Una neurona del hipocampo inicialmente respondiendo solo a Jennifer Aniston comienza a responder a la Torre Eiffel cuando se genera la asociación entre estos dos conceptos.

Estos resultados me han llevado a concluir (uso la primera persona para aclarar que esto es una interpretación personal y no algo universalmente aceptado) que las neuronas de concepto codifican asociaciones, porque el cimiento de la memoria es justamente hacer asociaciones entre conceptos. Mi recuerdo de aquel día en la playa con mi hermano se basa en asociar los conceptos involucrados (hermano,

playa, Brasil). Aunque evidentemente la memoria también involucra muchos otros aspectos. Recuerdo los rasgos de la cara de mi hermano, el sonido de las olas y el olor de las algas en la costa. Estos recuerdos son mucho más que abstracciones o asociaciones entre conceptos. Pero la representación de detalles se da en la corteza cerebral, particularmente en las áreas involucradas en el procesamiento de la información con los distintos sentidos (los detalles de una cara estarán en la corteza visual, los de un sonido en la corteza auditiva, etcétera). La codificación de detalles en la corteza está ligada a la representación de conceptos en el hipocampo, lo cual nos permite relacionar distintas impresiones sensoriales (el olor, la textura y el color de las algas marinas).

En el hipocampo tenemos una representación conceptual, de etiquetas, con la que es fácil generar nuevas asociaciones[210]. Si en vez de neuronas de concepto, tuviéramos una representación de detalles en el hipocampo, para cada nueva asociación, por ejemplo, la mi hermano con la playa de Brasil, haría falta establecer conexiones entre los detalles de estos conceptos, pero sin mezclarlos con otros, lo cual sería mucho más complicado ya que mi hermano se parece a otras personas que conozco y la playa en Brasil es parecida a otras playas del Mediterráneo.

En general, da la impresión de ser una simplificación desmesurada argumentar que la memoria de nuestras experiencias está basada en asociaciones entre conceptos. Aunque probablemente esto se deba a una noción errada que tenemos sobre la memoria, de creer que recordamos mucho más de lo que realmente podemos. Más allá de las representaciones guardadas en la corteza, muchos detalles que creemos recordar quizás no sean más que inferencias inconscientes que hacemos usando el sentido común. Tal vez en realidad no recuerde haberlo visto a mi hermano transpirar bajo el sol de febrero en Brasil, tirado en la arena como un lagarto, pero infiero que seguramente fue así y esta inferencia, que tiene mucho sentido, pasa a formar parte de mi memoria y la enriquecerá cuando evoque este recuerdo en el futuro.

Considerando toda la información que hemos observado, volvamos

a analizar el experimento de Akakhievitch. ¿Qué pasaría si destruyéramos cada una de las neuronas codificando a Jennifer Aniston en el hipocampo? ¿Acaso olvidaríamos completamente a la actriz? Esto recuerda a *Eterno resplandor de una mente sin recuerdos*, en donde a partir de un procedimiento similar, una pareja decide borrar completamente la memoria del otro tras una pelea dramática.

¿Qué pasaría si eliminamos todas las neuronas codificando el concepto Jennifer Aniston del hipocampo de una persona?[211]

A partir de la evidencia proporcionada por el paciente H. M., de quien habláramos anteriormente, sabemos el hipocampo es vital para la memoria; sin hipocampo no podemos formar nuevos recuerdos. Sin embargo, H. M. recordaba cosas del pasado: sabía quién era su madre, sus amigos y el cirujano que le hizo la operación. Volviendo al ejemplo de las vacaciones con mi hermano, la memoria de mi

hermano y de Brasil (es decir, saber quién es mi hermano, dónde queda Brasil, etcétera) está en la corteza cerebral; lo que hace el hipocampo es asociar rápidamente estos conceptos para luego poder recordar la experiencia de aquel día en la playa. Si eliminara una a una las neuronas que codifican a mi hermano en mi hipocampo, olvidaría los detalles de las vacaciones y de otras tantas asociaciones que constituyen los recuerdos de las experiencias vividas con mi hermano, pero aún sabría quién es él[212].

Borrar selectivamente el recuerdo de mi hermano, de Jennifer Aniston o cualquier otra persona —es decir, que tras la cirugía no sepamos quién es— sería imposible, porque hasta donde sabemos en la corteza no hay neuronas de concepto representando selectivamente personas, objetos u otros conceptos específicos[213]. En la corteza la representación es distribuida —una democracia de millones— y los conceptos están mayormente codificados por los mismos grupos de neuronas que disparan de una u otra manera para diferenciarlos. Es decir, en la corteza no hay un grupo de neuronas que dispara específicamente a Jennifer Aniston y otro que responde a Maradona; no hay activaciones específicas que podamos eliminar para borrar la memoria de una u otra persona. Sin embargo, podríamos en principio manipular la representación de cada concepto en el hipocampo y así alterar las asociaciones que subyacen los recuerdos de experiencias vividas.

Me arremango frente al teclado porque estoy por describir uno de los resultados más espectaculares de los últimos años en neurociencia. En un artículo publicado en 1979, Francis Crick (a quien ya mencionamos) observó que prácticamente cualquier comportamiento que vayamos a estudiar involucra una interacción compleja de distintas áreas del cerebro y que, consecuentemente, sería invaluable disponer de un método que pudiera alterar selectivamente la actividad de grupos de neuronas específicos, sin modificar el resto. La idea subyacente es pasar de estudios de *correlación* a estudios de *causalidad*. Por ejemplo, la actividad de neuronas en una cierta área puede estar correlacionada con un comportamiento, pero no estar involucrada en la generación de este; no ser su causa. Veamos esto con un ejemplo.

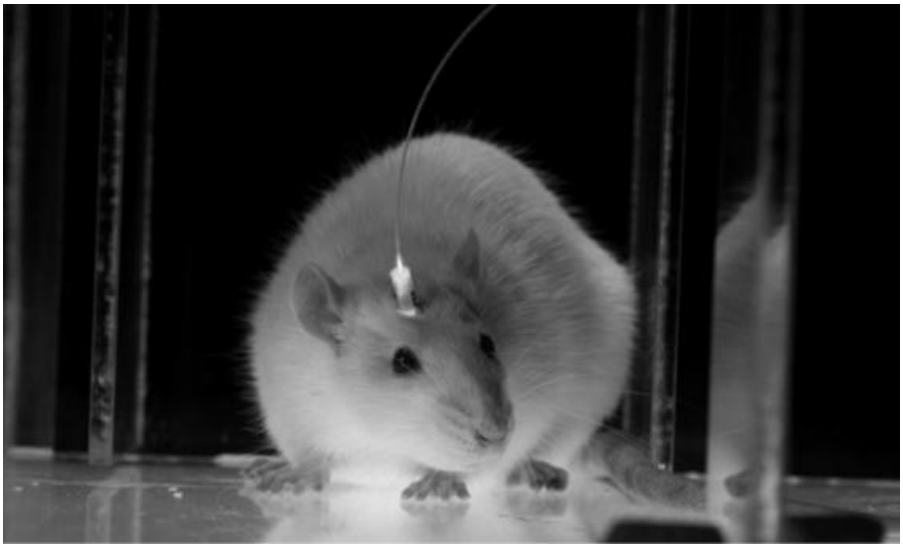
El salario promedio de una persona está correlacionado con el tamaño de su pie. Suena ridículo, pero es una verdad irrefutable. Una persona que calza 43 tiende a ganar más que otra que calza 39; en promedio, cuanto más grande el pie, mayor el salario. La razón de esta correlación tan bizarra es que el hombre tiende a ganar más que la mujer y suele tener el pie más grande. La relación con el tamaño del pie es una consecuencia secundaria, pero es real; podemos tomar millones de personas al azar y obtendremos una correlación altísima. Mas si quisiéramos demostrar que efectivamente es el tamaño del pie lo que determina el salario, podríamos buscar la manera de hacer crecer el pie de un grupo de personas (no sabría cómo hacerlo, pero imaginemos que fuera posible) y ver si eso afecta su sueldo a fin de mes. Este estudio, de causalidad, en el cual modificamos una variable y vemos cómo afecta la otra, daría evidencia mucho más contundente sobre la relación entre los dos efectos que una simple correlación.

Volviendo entonces a la afirmación de Crick, supongamos que tenemos neuronas en algún área del cerebro cuya actividad está correlacionada con un determinado comportamiento, por ejemplo, leer. Pero es posible que estas neuronas no respondan al hecho de leer en sí mismo, sino a un mayor grado de atención que requiere la lectura; es decir, si hiciéramos cualquier tarea que implique prestar atención veríamos lo mismo. Idealmente, querríamos hacer un estudio de causalidad y desactivar temporariamente, como con un interruptor, la actividad de estas neuronas y ver si es cierto que cuando las apagamos no podemos leer y que cuando las prendemos podemos leer de nuevo. Crick luego especularía que un potencial método para controlar el disparo de las neuronas debería involucrar activaciones a través de la luz, ya que esta puede prenderse y apagarse con gran precisión temporal. Décadas más tarde se mostró que no estaba tan errado.

Ciertas algas tienen receptores de luz que les permiten, usando pequeños filamentos, nadar en dirección a la luz para hacer fotosíntesis como el resto de las plantas. Esto lo hacen a través de proteínas que funcionan como células fotovoltaicas, las cuales convierten la luz en minúsculas descargas eléctricas para guiar el

movimiento. Dejando de lado una infinidad de detalles técnicos, la gran revolución de los últimos años en neurociencia fue implantar estas proteínas en las neuronas de distintos animales, activando o inhibiendo su actividad al iluminarlas[214]. Para hacer esto, la complicación radica, por un lado, en efectivamente lograr introducir estas proteínas en las neuronas y, por el otro, en iluminar dentro del cerebro para manipular su activación. Lo primero se logró haciendo uso de avances en ingeniería genética, "infectando" las neuronas con un virus que en su ADN contenía la proteína que reaccionaba a la luz de las algas, y lo segundo se hizo a través de pequeños cables de fibra óptica implantados en el cerebro[215].

Hablamos de Penfield y de cómo lograba evocar reacciones diversas estimulando eléctricamente distintas áreas del cerebro. Pero la estimulación a través de la luz, lo que llamamos *optogenética*, tiene muchísimas ventajas, ya que pueden estimularse grupos de neuronas específicas (sin alterar las otras neuronas, algo muy difícil de hacer con estimulación eléctrica ya que, a no ser que se coloquen electrodos tocando cada neurona, se estimulan indiscriminadamente todas las neuronas alrededor del electrodo usado) y el control de la activación o desactivación de neuronas es mucho más preciso. Hace poco más de 10 años y casi 30 después de haber propuesto controlar la actividad de las neuronas con solo iluminarlas, "prendiéndolas" y "apagándolas" con un interruptor de luz, el sueño de Francis Crick se volvía realidad.



Optogenética: a partir de iluminar dentro del cerebro usando una fibra óptica, se pueden activar o inhibir a neuronas genéticamente manipuladas para reaccionar a la luz[216].

Estudios iniciales efectivamente mostraron que a partir de la optogenética era posible cambiar la actividad de las neuronas[217], pero la gran pregunta era si tales manipulaciones podrían llegar a modificar el comportamiento de los animales, mostrando un rol causal y no solo una correlación entre la activación de las neuronas y las distintas funciones del cerebro.

Los primeros resultados tardaron un par de años en llegar, pero fueron contundentes. Un grupo de neuronas en el hipotálamo en ratones (que había sido implicada en la regulación de períodos de sueño y vigilia y también con la narcolepsia) fue estimulado usando optogenética, haciendo que los ratones despertaran inmediatamente de su sueño[218]. En otro caso se usó optogenética para estimular neuronas en la corteza motora de ratones, logrando que estos movieran sus bigotes, y estimulando un área motora relacionada a la locomoción; otras pruebas mostraron cómo los ratones comenzaban a correr alocadamente cada vez que se iluminaban las neuronas de esta área[219]. Pero, así como manipulamos los movimientos o los períodos de sueño y vigilia, ¿podremos también manipular los recuerdos?

En pacientes con los que hicimos registros en el hipocampo, vimos que al formar nuevos recuerdos se pudo cambiar la actividad de las

neuronas codificando los conceptos involucrados. La gran pregunta es si podremos hacer lo contrario: manipular la actividad de neuronas en el hipocampo usando optogenética y así crear nuevos recuerdos. Esto es muy distinto a crear falsas memorias a partir de la sugestión y los procesos de reconsolidación. Ahora nos planteamos algo mucho más espectacular; estamos hablando de implantar recuerdos a partir de estimular directamente el cerebro.

Susumu Tonegawa obtuvo el Premio Nobel en 1987 por su descubrimiento de los mecanismos genéticos que producen la diversidad de anticuerpos. Pero poco después de haber recibido el Nobel cambió radicalmente su campo de investigación: de inmunología a neurociencia. Desde entonces, su laboratorio ha venido reportando una serie de descubrimientos espectaculares, particularmente usando ratones transgénicos —ratones con mutaciones genéticas que producen alteraciones en la actividad de neuronas y las conexiones entre estas— para dilucidar el rol de las distintas áreas que constituyen el hipocampo en la formación de la memoria.

Los resultados más impresionantes ocurrieron cuando Tonegawa y su equipo hicieron uso de la optogenética. En un primer estudio crearon un reflejo condicionado en ratones —la presentación de un sonido era seguida por una molesta descarga eléctrica— lo cual hacía que los ratones se quedaran totalmente inmóviles al escuchar el sonido, previniendo la descarga. Pero llevaron a cabo el aprendizaje del reflejo condicionado usando una manipulación farmacológica/genética muy ingeniosa. Los ratones recibían un compuesto bloqueando la incorporación del virus con la proteína fotosensitiva de las algas y los investigadores dejaron de darles este compuesto mientras aprendían el reflejo condicionado. Al hacerlo, infectaron las neuronas relacionadas con este reflejo —aquellas que estaban activas cuando se produjo la infección con el virus— y por lo tanto la estimulación optogenética producida al iluminar dentro del cerebro (más específicamente, en un área del hipocampo llamada *giro dentado*) activaba solo estas neuronas.

El contexto espacial es crucial para los ratones; saber cómo

comportarse en uno u otro ambiente es una cuestión de vida o muerte, reconociendo si pueden moverse libremente en una guarida segura o si tienen que andar con cuidado por la amenaza de un gato u otro predador. Volviendo al experimento de Tonegawa, lo que hicieron fue generar el reflejo condicionado en un ambiente dado y luego colocaron a los ratones un ambiente distinto, en donde nunca antes habían recibido una descarga eléctrica. Los ratones se comportaban normalmente en este nuevo ambiente "seguro", pero al prender la luz y activar las neuronas involucradas en la codificación del reflejo condicionado realizado en el primer ambiente, quedaron inmóviles, como esperando una descarga. O sea, al activar neuronas específicas lograron evocar una memoria y el comportamiento asociado a esta[220].

Aún más espectacular fue la siguiente prueba. Primero usaron el mismo tipo de manipulación farmacológica/genética para infectar con el virus conteniendo la proteína fotosensitiva a las neuronas activas mientras los ratones exploraban un cierto ambiente. Luego llevaron a los ratones a otro ambiente donde, como antes, crearon un reflejo condicionado (la presentación de un sonido era seguida por una descarga eléctrica), pero mientras los ratones aprendían este reflejo prendieron el haz de luz para activar las neuronas que representaban el primer ambiente, creando artificialmente la asociación entre ese primer ambiente neutro y la aparición de la descarga eléctrica. Finalmente volvieron a poner a los ratones en el primer ambiente, donde no había ocurrido absolutamente nada, y observaron que se comportaban como en el ambiente en donde habían aprendido el reflejo condicionado, mostrando el mismo tipo de inmovilidad que tenían al predecir la descarga eléctrica. En otras palabras, a partir de activar grupos específicos de neuronas lograron generar una memoria falsa: relacionar un lugar neutro y seguro con descargas eléctricas que ocurrieron en otro lado[221].

Volvamos a *Total Recall* y a cómo eventualmente podríamos transformar a Hauser en Quaid. Para ello deberíamos, por un lado, borrarle los recuerdos que lo ligan a Marte y su rol como agente de la gran corporación que domina el planeta, y por el otro, implantarle memorias falsas, como la de ser un obrero trabajando en una fábrica y

la de estar casado casado con su mujer. En conclusión, deberíamos borrar algunas asociaciones e implantar otras, por ejemplo usando optogenética. Lo interesante es que el implante de falsas memorias se basaría en generar asociaciones; en principio no sería necesario implantar los recuerdos en detalle, ya que de eso se encarga naturalmente el cerebro, generando una construcción alrededor de pocas asociaciones entre conceptos y enriqueciendo la memoria con cada evocación y reconsolidación, así como mostrara Loftus al implantar el falso recuerdo del centro comercial. Todavía suena a ciencia ficción, aunque es algo en principio posible y quizás no tan lejano.

Pero ¿podremos usar optogenética en humanos, podremos manipular a voluntad la actividad de grupos específicos de neuronas y los comportamientos asociados con estas, como ya se mostró en ratones? La optogenética implica, por un lado, usar un virus para hacer fotosensitivas a las neuronas y, por el otro, la implantación de diminutas fibras ópticas dentro del cráneo para iluminarlas. Aparentemente, esto es algo que dista mucho de ser éticamente aceptable, pese a que ya existen muchos casos en donde tales intervenciones tendrían sentido. Trabajos recientes empiezan a mostrar cómo manipulaciones a través de la optogenética ofrecen alternativas terapéuticas muy promisorias para tratar, por ahora con modelos en animales, la epilepsia, la esquizofrenia, trastornos de ansiedad, adicciones, depresión y Parkinson, entre otras patologías[222]. No sería sorprendente, por lo tanto, que los avances demostrados en animales empiecen en breve a ser aplicados a humanos.

Volvamos una vez más a Akakhievitch e imaginemos que su nieto Drogori, tras leer sobre los avances y los grandes desafíos de la neurociencia en estos últimos años, decide limpiar el nombre de su abuelo, tan desprestigiado a partir de la historia de Lettvin. Casi nadie toma en serio la idea de las “neuronas de la abuela”, ya que ni siquiera las neuronas de concepto en el hipocampo están cerca de hacer lo que dijera su abuelo Akakhi. Pero Drogori concibe una idea más brillante y diabólica. Drogori ya sabe de los experimentos de Loftus y Tonegawa; sabe de la posibilidad de borrar e implantar

memorias, pero lo que realmente le fascina es entender qué es lo que distingue a los humanos del resto de los animales.

Los avances de la Inteligencia Artificial en estos últimos años nos llevan a replantearnos qué define nuestra inteligencia y si alguna vez crearemos seres de silicio que lleguen a superarnos. Pero las computadoras todavía carecen de una inteligencia general; no pueden abstraer, extraer un sentido de las cosas, inferir y hacer analogías. Las computadoras almacenan fehacientemente todo y no seleccionan y elaboran la información que es esencial. Sin embargo, no sería sorprendente si en unos años algún genio de la informática encontrara la manera de hacer que las computadoras puedan empezar a entender el contenido de lo que almacenan, replicando los procesos que realiza el cerebro humano. Por ende, lo que a Drogori realmente le intriga es la diferencia entre los humanos y el resto de los animales; en particular, con los primates superiores. Después de todo, la estructura del cerebro es muy parecida, solo que los monos carecen de neuronas de concepto; o para decirlo de una manera algo más temeraria, la existencia de neuronas de concepto probablemente sea la gran diferencia cualitativa entre el cerebro del mono y el humano[223].

Fascinado por la historia de *El planeta de los simios*, en donde un mono, Cesar, comienza a pensar y hablar como un humano, Drogori concibe un experimento singular para implantar neuronas de concepto en monos. Sabe que la falta de estas neuronas en animales probablemente se deba al hecho de que no tienen lenguaje, que es lo que en el fondo reafirma las representaciones de conceptos; pero Drogori tiene en mente un atajo para lograrlo, usando la optogenética y experimentos inspirados en aquellos hechos por Tonegawa. Su idea es activar selectivamente un grupo específico de neuronas en el hipocampo del mono mientras le muestra fotos de, por ejemplo, uno de los veterinarios a su cargo. La activación generará un engrama, un grupo de neuronas que responderá al veterinario. Pero la clave radica en no mostrarle una única foto, sino una gran variedad de fotos distintas del veterinario —de frente, de perfil, con el ambo del laboratorio, vestido de sport, de traje, etcétera— para así crear una representación del concepto y no de alguna foto específica; es decir,

inducir neuronas que respondan al veterinario, sin importar cómo se vea. Más aún, Drogori planea no solo usar imágenes sino también estímulos auditivos, como el sonido de la voz del veterinario, e incluso olfativos, su olor en particular, a fin de generar una representación multimodal que combine las sensaciones de los distintos sentidos.

El resultado de Drogori reflotaría la vieja idea de su abuelo —generaría neuronas que responden a una persona específica— pero su objetivo es mucho más ambicioso. Así como podría crear un engrama representando al veterinario, también podría crear otros representando a los otros monos de la colonia, a las personas con las que el mono tiene contacto, distintos tipos de comida y demás conceptos que estén presente en su día a día. Básicamente, Drogori estaría creando una representación explícita de conceptos en el hipocampo, que el mono eventualmente podría empezar a usar para formar nuevas asociaciones y memorias episódicas. Es decir, generaría la representación abstracta de las cosas que usamos en nuestro pensamiento. La gran pregunta que se hace Drogori, entonces, es si al crear estas representaciones —algo en principio factible, pues no es más que un experimento análogo a los que ya se hicieron con ratones — podría lograr que el mono piense y llegue a desarrollar la inteligencia de un humano. En un libro sobre ciencia y ciencia ficción, me permito este argumento ficticio y algo temerario (aunque posible) para cerrar este capítulo.

188 La película fue estrenada en 1990 (con una reedición en 2012) y está basada en un cuento (cuándo no) de Philip Dick llamado “We can remember it for you wholesale”.

189 Héctor Maldonado, pionero de la neurociencia y del estudio de la memoria en la Argentina, argumenta que no es casual que Mnemósine sea la madre de las musas, ya que de la memoria surge la inspiración artística (Héctor Maldonado; *La memoria animal*; Eudeba; 2008).

190 La imagen de la izquierda corresponde a una tabla de cera del antiguo Egipto expuesta en el Metropolitan Museum of Art, y la ilustración de la derecha fue publicada en *National Geographic*.

191 H. Ebbinghaus; *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*; Duncker & Humblot; Leipzig; 1885.

192 James McGaugh. "Memory –a century of consolidation". *Science* 287: 248-251; 2000.

193 James Misanin, Ralph Miller and Donald Lewis. "Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace". *Science* 160: 554-555; 1968.

194 Karim Nader, Glenn Schafe and Joseph Le Doux. "Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval". *Nature* 406: 722-726; 2000.

195 Ver por ejemplo:

- Karim Nader. "Memory traces unbound". *Trends in Neurosciences* 26: 65-72; 2003.
- Karim Nader and Oliver Hardt. "A single standard for memory: the case for reconsolidation". *Nature Reviews Neuroscience* 10: 224-234; 2009.

196 Merel Kindt, Marieke Soeter and Bram Vervliet. "Beyond extinction: erasing human fear responses and preventing the return of fear". *Nature Neuroscience* 12: 256-258; 2009.

197 Para una revision de estos resultados ver: Tom Beckers and Merel Kindt. "Memory reconsolidation interference as an emerging treatment for emotional disorders: Strengths, limitations, challenges and opportunities". *Annual Reviews of Clinical Psychology* 13: 99-121; 2017.

198 E. Loftus and J. Palmer. "Reconstruction of automobile destruction: an example of interaction between language and memory". *Journal of Verbal Learning and Verbal Behavior* 13(5): 585-589; 1974.

199 Estos hallazgos tuvieron gran repercusión ya que pusieron de manifiesto la subjetividad de los testigos visuales durante juicios orales y el modo en que un testimonio puede ser manipulado según la forma en que se formulen las preguntas.
Para más detalles, ver el capítulo 4 de *Qué es la memoria*, op. cit., y: Elizabeth Loftus. "Our changeable memories: legal and practical implications". *Nature Reviews Neuroscience* 4: 231-234; 2003.
Por otro lado, la maleabilidad de los recuerdos durante su reconsolidación nos alerta sobre el poder de la propaganda política y la repetición e interpretación de noticias en los medios. Esto está muy ligado a la discusión que hiciéramos al final del capítulo 6 sobre la ceguera de elección y lo fácil que es manipular decisiones.

200 Para más detalles, ver:

- Elizabeth Loftus. "Creating false memories". *Scientific American* 277: 70-75; 1997.
- Elizabeth Loftus and Jacqueline Pickrell. "The formation of false memories". *Psychiatry Annals* 25: 720-725; 1995.

201 Para más detalles, ver: Rodrigo Quian Quiroga. "Magic and cognitive neuroscience". Op. cit.

202 Para más detalles, ver:

- Rodrigo Quian Quiroga, Itzhak Fried and Christof Koch. “Brain cells for grandmother”. *Scientific American* 308: 30-35; 2013.
- Charles Gross. “Genealogy of the ‘Grandmother Cell’”. *The Neuroscientist* 8: 512-518; 2002.
- Horace Barlow; “The neuron doctrine in perception” en *The Cognitive Neurosciences*; MIT Press (Boston); 1994.

203 William James; *The Principles of Psychology*; vol. I; capítulo 6; 1890.
Nociones parecidas fueron propuestas por Santiago Ramón y Cajal, quien hablara de neuronas psíquicas en la corteza, responsables funciones mentales complejas, y por Sigmund Freud, quien postulara la existencia de neuronas Psy y neuronas Phi, encargadas respectivamente de la percepción y la memoria.
204 Charles Sherrington; *Man on his Nature*; Cambridge University Press; 1940. Tiempo después, el neurocientifico británico Horace Barlow argumentaría que el cerebro no toma decisiones en base a un solo *pontífice* o una *democracia de millones*, sino usando un *colegio de cardenales*; es decir, el cerebro no usa neuronas únicas ni redes distribuidas, sino una representación con pocas neuronas (pero no una sola). Esto es lo que en neurociencia se conoce como una representación *sparse* (poco densa, rala). Para más detalles, ver: Horace Barlow. “Single units and sensation: A neuron doctrine for perceptual psychology?” *Perception* 1: 371-394; 1972.
205 Jerzy Konorski; *Integrative activity of the brain. An interdisciplinary approach*; University of Chicago Press; Chicago; 1967.
206 Op. cit.
Las respuestas neuronales en V1, en IT y en el hipocampo fueron publicadas en:

- David Hubel and Torsten Wiesel. “Receptive fields of single neurones in the cat’s striate cortex”. *Journal of Physiology* 148: 574–591; 1959.
- Charles Gross. “Single neuron studies of inferior temporal cortex”. *Neuropsychologia* 46: 841–852; 2008.
- Rodrigo Quian Quiroga, Alexander Kraskov, Christof Koch and Itzhak Fried. “Explicit encoding of multimodal percepts by single neurons in the human brain”. Op. cit.

207 Para un estudio cuantitativo de esta afirmación, ver: Emanuela de Falco, Matias Ison, Itzhak Fried and Rodrigo Quian Quiroga. “Long-term coding of personal and universal associations underlying the memory web in the human brain”. *Nature Communications* 7: 13408; 2016.
208 Para más detalles, ver: Rodrigo Quian Quiroga. “Concept cells: the building blocks of declarative memory functions”. *Nature Reviews Neuroscience* 13: 587-597; 2012.
209 Matias Ison, Rodrigo Quian Quiroga and Itzhak Fried. “Rapid encoding of new memories by individual neurons in the human brain”. *Neuron* 87: 220-230; 2015.
210 Rodrigo Quian Quiroga. “Concept cells: the building blocks of declarative memory functions”. Op. cit.

211 Esta figura fue hecha por Dave Barton para ilustrar una charla que alguna vez diera sobre la neurona de Jennifer Aniston en un Café Scientific en Inglaterra.

212 Cabe aclarar que esta distinción no es tan tajante, pues algunas asociaciones en principio *episódicas* (establecidas en el hipocampo) también pueden pasar a formar parte de la *memoria semántica* guardada en la corteza cerebral. Por ejemplo, puedo recordar el evento de ver por primera vez a mi hermano con su pareja, lo cual es una memoria episódica, pero tras otros tantos encuentros generaré la memoria semántica, fáctica, de saber quién es la pareja de mi hermano, sin necesariamente recordar aquel primer encuentro (aunque en este caso, tengo muy presente distintas asociaciones de esa noche memorable).

La discusión sobre cuáles memorias y asociaciones se transfieren a la corteza cerebral en el proceso de consolidación, y cuáles otras quedan guardadas en el hipocampo, ha dado lugar a acalorados debates en la neurociencia, resultando en dos modelos antagónicos: el Standard Consolidation Model afirma que tanto las memorias episódicas como las semánticas se consolidan en la corteza y que el hipocampo solo tiene una función temporaria en la formación de la memoria, mientras que el Memory Trace Theory manifiesta que las memorias episódicas quedan codificadas permanentemente en el hipocampo.

Para más detalles, ver: Morris Moscovitch *et al.* “Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory”. *Journal of Anatomy*. 207: 35–66; 2005.

213 Aclaro que hablo de borrar selectivamente un concepto dejando intacto los otros. Experimentos en ratas mostraron que la memoria también puede borrarse, aunque sin tal especificidad, inhibiendo una proteína llamada PKMζ, la cual es fundamental para mantener la conectividad entre neuronas (representando memorias).

Para más detalles, ver:

- Eva Pastalkova, Peter Serrano, Deana Pinkhasova, Emma Wallace, Andre Fenton and Todd Sacktor. “Storage of spatial information by the maintenance mechanism of LTP”. *Science* 313: 1141-1144; 2006.
- Reut Shema, Todd Sacktor, Yadin Dudai. “Rapid erasure of Long-Term Memory associations in the cortex by an inhibitor of PKMζ”. *Science* 317: 951-953; 2007.

214 La activación de las neuronas se obtiene usando una proteína llamada *channelrhodopsin*, mientras que para la inhibición se usa otra denominada *halorhodopsin*. En la actualidad se han desarrollado proteínas que tienen distintos efectos en la actividad de las neuronas, las cuales son activadas con luz de distinto color (azul, amarillo, verde). Suena realmente increíble, pero luego de haber realizado el proceso genético de introducir las proteínas en las neuronas, pueden inducirse distintos tipos de manipulaciones en su actividad, según sea el color de la luz usada.

215 Para más detalles, ver:

- Karl Deisseroth. “Optogenetics: Controlling the brain with light”. *Scientific American* 303: 48-55; 2010.
- Karl Deisseroth. “Optogenetics: 10 years of microbial opsins in neuroscience”. *Nature Neuroscience* 18: 1213-1225; 2015.

216 Crédito de la imagen: John Carnett, Viviana Gradinaru, Murtaza Mogru y Karl Deisseroth de la Universidad de Stanford.

217 E. Boyden, F. Zhang, E. Bamberg, G. Nagel and K. Deisseroth. “Millisecond-timescale, genetically targeted optical control of neural activity”. *Nature Neuroscience* 8: 1263-1268; 2005.

218 Antoine Adamantidis, Feng Zhang, Alexander Aravanis, Karl Deisseroth and Luis de Lecea. “Neural substrates of awakening probed with optogenetic control of hypocretin neurons”. *Nature* 450: 420-424; 2007.

219 El experimento logrando mover los bigotes de los ratones se publicó en: Alexander Aravanis, Li-Ping Wang, Feng Zhang, Leslie Meltzer, Murtaza Mogri, M. Bret Schneider and Karl Deisseroth. “An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology”. *Journal of Neural Engineering* 4: S143-S156; 2007.

Y para una demostración de cómo la estimulación de un área motora hacía que el ratón corra alocadamente ver: https://www.youtube.com/watch?v=88TVQZUfYGw

220 Xu Liu, Steve Ramirez, Petti Pang, Corey Puryear, Arvind Govindarajan, Karl Deisseroth and Susumu Tonegawa. “Optogenetic stimulation of a hippocampal engram activates fear memory recall”. *Nature* 484: 381-385; 2012.

221 Steve Ramirez, Xu Liu, Pei-Ann Lin, Junghyup Suh, Michele Pignatelli, Roger Redondo, Tomas Ryan and Susumu Tonegawa. “Creating a false memory in the hippocampus”. *Science* 341: 387-391; 2013.

222 Para una revision de estos resultados, ver: Kay Tye and Karl Deisseroth. “Optogenetic investigation of neural circuits underlying brain disease in animal models”. *Nature Reviews Neuroscience* 13: 251-266; 2012.

223 En el capítulo 3 discutimos la similitud entre las *place cells*, encontradas en el hipocampo de ratones, y las neuronas de concepto en el hipocampo humano. Sin embargo, las *place cells* responden de distinta manera si se cambia el contexto y tienen por lo tanto una muy limitada capacidad de abstracción, en contraste con las neuronas de concepto, que responden de la misma manera en distintos experimentos y contextos. En monos, lo más cercano a estas neuronas son las neuronas con respuestas a caras en los *face patches*, mencionadas en el capítulo 5. Sin embargo, recientemente se mostró que estas neuronas no responden a personas específicas sino a detalles visuales de las distintas caras. Para más detalles, ver:

- Steve Chang and Doris Tsao. “The Code for Facial Identity in the Primate Brain”. Op. cit.
- Rodrigo Quian Quiroga. “How do we recognize a face?” Op. cit.

Recientemente un grupo de investigadores también replicó, haciendo registros en el hipocampo de monos, el experimento con el que obtuvimos neuronas de concepto en

humanos, pero sin encontrar este tipo de neuronas. Para más detalles, ver: Julia Sliwa, Aurélie Planté, Jean-René Duhamel and Sylvia Wirth. “Independent neuronal representation of facial and vocal identity in the monkey hippocampus and inferotemporal cortex”. *Cerebral Cortex* 26: 950-966; 2014.

X. Abre los ojos

Inmortalidad

Muchas generaciones de argentinos crecimos leyendo historias de personajes extraordinarios en las revistas de la editorial Columba: *El Tony*, *Intervalo*, *Fantasía*, *D'Artagnan*, *Nippur Magnum*[224]. En mayo de 1969, *D'Artagnan* lanzaría "Gilgamesh el inmortal", una historieta hecha por el dibujante y guionista Lucho Olivera, que inicialmente iba a tener un único capítulo, pero terminó convirtiéndose en una serie con más de 140 entregas por pedido de los lectores.

La historia está basada en una bellísima leyenda sumeria, en la cual Gilgamesh, rey de Uruk más de dos milenios antes de Cristo, emprende un viaje a través de desiertos, mares y montañas para encontrar el secreto de la vida eterna. Después de una larga travesía, encuentra a Utnapishtim, el Lejano, el único hombre al que los dioses concedieron la inmortalidad, quien le revela que el secreto de la vida eterna está en una planta de licio en el fondo del mar. Gilgamesh ata piedras a sus pies para llegar al fondo de las aguas y una vez que encuentra la planta, se desprende de las piedras y regresa a su barca con el preciado elixir. Sin embargo, una serpiente le roba la planta de licio en su viaje de regreso, despojándolo del don de la inmortalidad (aunque según la leyenda, Gilgamesh reinó Uruk durante 126 años).

Nuestra vida pasa muy rápido. Vivimos unas cuantas décadas y tarde o temprano enfrentamos el terrible fantasma de la muerte. Mas los avances extraordinarios de la ciencia llegan incluso a desafiar la inefabilidad de nuestra extinción y, así como Gilgamesh en la antigua Uruk, buscamos aquella flor de licio que nos lleve a perpetuarnos por el resto de los tiempos. Es el impulso más fuerte de todo ser vivo, el *conatus*, ese instinto de supervivencia arraigado en cada uno de nosotros que nos hace temer la muerte y aferrarnos a la vida.

El dualismo cartesiano, advocado por las religiones occidentales, ofrece un grato consuelo: con la muerte perece el cuerpo, pero la mente, aquello que define nuestro ser, sobrevive la extinción de la

materia y nos hace inmortales. Mas como viéramos en el capítulo 1, el dualismo cartesiano es difícil de sostener desde el punto de vista filosófico. Aunque vale aclarar que la ciencia, la razón, no puede demostrar ni la existencia ni la inexistencia de un Dios o de una vida más allá de la muerte; eso es una cuestión de fe. Pero dejemos de lado argumentos religiosos y analicemos las posibilidades ciertas que ofrece la ciencia de trascender.

Los científicos de hoy mayormente abrazamos la idea del materialismo: no existe la mente separada del cuerpo, son una misma cosa. Hay entonces un sustrato material —la actividad de las neuronas — que determina nuestra memoria e identidad. Y es justamente esta afirmación la que abre la posibilidad de lograr lo aparentemente imposible y vencer a la muerte en la gran partida de ajedrez, como retrata Ingmar Bergman en *El séptimo sello*.

Embebidos en esta, la más formidable cruzada científica de todos los tiempos, imaginemos entonces que podamos mantener o, más específicamente, crionizar nuestro cuerpo al perecer; es decir, conservarlo a muy baja temperatura evitando la descomposición de los órganos, para eventualmente resucitarlo cuando los avances de la ciencia lo permitan, en decenas o tal vez cientos de años. Después de todo, ¿quién podría asegurar que esto no será posible en el futuro? Hasta hace unos pocos años también parecía imposible crear cyborgs o computadoras que derrotaran a campeones de ajedrez y, un poco más atrás en el tiempo, también era una quimera que el hombre llegara a la Luna.

La noción de la crionización ha sido usada en infinidad de películas de ciencia ficción[225]. Entre ellas *Abre los ojos*, de Alejandro Amenábar, en la que una persona es preservada en un continuo sueño virtual esperando ser resucitada[226]. Hasta aquí, nada nuevo —la película de Amenábar es otra de tantas usando el mismo recurso argumental; pero con una gran diferencia: *Abre los ojos* esboza cómo sería el sueño mantenido durante décadas, dando lugar a una trama exquisitamente compleja.

Antes nos preguntamos si quizás un genio maléfico, el famoso genio de Descartes, podría generarnos la ilusión de una realidad inexistente

a partir de estimular nuestros sentidos. Una idea semejante es tratada en *Matrix*, en donde se combinan dos realidades paralelas: la de *Matrix*, con humanos en incubadoras viviendo una realidad inexistente, y la de aquellos que se liberan de esta realidad para enfrentar a las máquinas que la generan.

La imagen que ilustra este capítulo es una de las primeras escenas de *Abre los ojos*, donde el personaje principal se encuentra absolutamente solo en la Gran Vía de Madrid. La soledad del personaje, en una ciudad típicamente atiborrada de gente, llama notablemente la atención y evoca los argumentos de Dan Dennett, quien alegara que la idea de *Matrix,* o del demonio de Descartes, es en realidad imposible, ya que no podemos replicar el infinitamente complejo *feedback* que tenemos al interactuar con otras personas y el mundo exterior. La escena en la Gran Vía finalmente no es más que un sueño, y una vez que despierta, la situación se repite en el cotidiano caos de todos los días. Estos cambios son quizás lo más atrapante de la película, y cómo en el relato se fusionan magistralmente la realidad y la imaginación, la vigilia y el sueño; cómo se mezclan las memorias y hasta la identidad de las personas para aquel que vive un sueño eterno. La complejidad del mundo exterior y nuestras interacciones con este nos demuestran entonces que la realidad exterior existe, mientras que la coherencia de los hechos que nos acontecen reafirman que no estamos soñando; por más elaborado que sea el sueño en el cual estuviéramos embebidos, la ilusión no puede prolongarse eternamente, en algún momento la narrativa comenzaría a fallar, así como sucede en *Abre los ojos*.

La trama refiere a una compañía de crionización en Arizona llamada Life Extension. Curiosamente, esta compañía existe. Su nombre es Alcor Life Extension Foundation y, como en la película, ofrece el servicio de preservar el cuerpo en nitrógeno líquido tras la muerte, permitiendo la posibilidad de restablecer la conciencia de la persona cuando la tecnología lo permita en el futuro[227].

El proceso es mucho más complejo que meramente bajar la temperatura hasta congelar el cuerpo. Funciona así: una vez declarada muerta la persona[228], se la coloca en un baño de hielo para reducir

rápidamente su temperatura corporal y se le aplica respiración asistida para evitar daños causados por falta de irrigación sanguínea, particularmente al cerebro; luego se inyecta un líquido anticoagulante en el tracto sanguíneo y finalmente se reemplaza la sangre del sistema circulatorio por un compuesto químico que, al ser llevado a temperaturas más bajas, va transformándose en un líquido viscoso hasta llegar a un estado vidrioso. Este proceso, llamado *vitrificación*, se lleva a cabo para evitar cristalizaciones de agua congelada que puedan romper los tejidos. Finalmente, el cuerpo es conservado a 196 grados bajo cero (la temperatura del nitrógeno líquido usado en este proceso), evitando cualquier tipo de reacciones químicas y garantizando su preservación durante siglos[230].

Tanque de crionización de hasta cuatro cuerpos de la compañía Alcor en Arizona[229].

Cabe aclarar que un cerebro crionizado en realidad no puede soñar, como en *Abre los ojos,* porque las neuronas permanecen congeladas e inactivas. Vimos que el sueño es mucho más que un período de descanso, y que posee un rol muy activo y primordial en el aprendizaje y, particularmente, en la consolidación de la memoria. Durante el sueño las neuronas no "duermen", sino que tienen una gran actividad, la cual genera todas las historias fantásticas que apenas recordamos al despertar. Por lo tanto, en vez de crionizar el cerebro

deberíamos mantenerlo vivo (y activo) para que la persona pudiera perpetuarse en el sueño[231]. En la crionización no podríamos soñar, pues estaríamos clínicamente muertos, sin la menor actividad cerebral.

Dejemos entonces de lado la posibilidad del sueño placentero hasta ser resucitados y tratemos una pregunta más primordial: ¿tiene sentido la idea de revivir a una persona congelada? En principio, la crionización parece algo disparatado y es considerada por muchos como una pseudociencia. Sin embargo, se basa en cierta evidencia científica y en técnicas muy usadas para trasplantes. De hecho, diversos órganos son rutinariamente congelados y preservados a bajas temperaturas, y cabe entonces preguntarse si no podríamos hacer lo mismo con el cerebro y la identidad de la persona.

En la década del 50, se mostró que era posible revivir ratas llevadas a una temperatura de 0 grado Celsius, las cuales permanecieron por más de una hora sin respiración ni ritmo cardíaco[232]. Mas el proceso de congelado a la temperatura del nitrógeno líquido puede que dañe las estructuras del cerebro. El problema es que los compuestos criogénicos son altamente tóxicos y, además, al usarlos con organismos complejos, se dificulta la posibilidad de controlar su difusión en los diversos tejidos. Más aún, la experiencia en conservación de órganos para trasplantes indica que el proceso varía según el órgano a preservar, utilizando distintas velocidades de congelamiento y diferentes concentraciones de líquidos criogénicos, por lo que el proceso de vitrificación de un cuerpo no es nada trivial.

Sin embargo, a partir de desarrollar una nueva técnica de crionización, investigadores de California recientemente preservaron el cerebro de un conejo a 135 grados bajo cero y al descongelarlo, analizando cortes del cerebro en el microscopio, mostraron que las neuronas y las conexiones entre estas no habían sido dañadas[233]. Si bien es un resultado alentador para los defensores de la crionización, aún quedan por resolver muchos problemas. En particular, es muy distinto analizar cortes del cerebro de un conejo que volverlo a la vida y comprobar que se comporta como antes de ser crionizado. Es más, no tenemos ni las más remota idea de cómo podríamos revivir a un

animal o a una persona tras la muerte (lo cual es muy distinto a congelar a un animal por una hora). Si supiéramos cómo llevarlo a cabo, ya lo estaríamos implementando en los hospitales; no haría falta congelar a las personas y esperar los avances de la ciencia en el futuro, las resucitaríamos ni bien mueren. (Y si en el futuro pudiéramos revivir a una persona de una muerte cerebral, deberíamos replantearnos qué definimos como *muerte*; en la sociedad capitalista en la que vivimos, no sería descabellado pensar que quizás terminaremos considerando muerta a una persona cuando no tiene suficiente dinero para costear los complejos tratamientos de resucitación.)

Los partidarios de la crionización, sin embargo, argumentan que por más que no tengamos idea de cómo resucitar a una persona, eso no quita que en el futuro podamos hacerlo. Después de todo, en el pasado se consideraba que una persona fallecía tras un paro cardíaco y hoy día puede reanimársela a través de procedimientos muy establecidos. Pero debe quedar claro que la crionización continúa siendo ciencia ficción, así como lo es la posibilidad de hacer viajes intergalácticos, utilizando deformaciones del espacio-tiempo. Lo que sí es dominio de la ciencia contemporánea es el proceso de crionización en sí mismo, de cómo llevar los tejidos del cuerpo a temperaturas tan bajas sin dañarlos. Esto dista mucho de ser algo probado y establecido. De hecho, no sería sorprendente que los científicos del futuro no puedan ni siquiera plantearse cómo revivir nuestros cuerpos debido a daños irreparables que hubieran sufrido los tejidos, y particularmente el cerebro, al ser congelados.

A pesar de todas estas críticas, para muchos es mejor tener una chance, por más ínfima que sea, a no tener absolutamente nada, y en nuestros días ya hay cientos de personas cuyos cuerpos (o solo sus cabezas) son preservados en tanques a bajas temperaturas, y otras miles ya registradas para recibir el tratamiento ni bien fallezcan[234]. Es una gran apuesta. Pero ¿quién puede arrojarse el don de predecir los avances de la ciencia? Aunque más allá de la posibilidad remota de resucitar un cuerpo crionizado, hay otros factores elementales que pueden interrumpir el proceso. Nada garantiza que los cuerpos sean

preservados por siglos porque podrían ser destruidos por un terremoto, una guerra nuclear o el simple hecho de que la compañía que presta los servicios haya quedado en bancarrota.

Vamos ahora a plantearnos una pregunta mucho más esencial, retomando la discusión sobre la identidad a partir de *Robocop*: ¿serán estas personas las mismas una vez resucitadas? ¿O será la crionización un procedimiento extremadamente caro y complejo para crear nuevas personas en el futuro?

Vimos que la noción de identidad está ligada al problema del cambio: las cosas cambian, pero continúan siendo las mismas; nosotros mismos cambiamos aunque seguimos siendo los de siempre. Pero ¿qué es lo que se mantiene, qué es lo que se pierde con la muerte?

Un día navideño en Suecia, una niña de 7 años dejó imprevistamente su casa para buscar a Papá Noel. Con una temperatura exterior de 12 grados bajo cero, cayó por un acantilado y quedó inconsciente, sumergida en aguas heladas. Luego de más de una hora fue rescatada, pero su situación era dramática: estaba en coma, con una temperatura corporal de 13 grados Celsius y sin signos vitales (no respiraba, ni tenía actividad cardíaca; sus pupilas estaban dilatadas y no respondían a la luz). Pero sorprendentemente, después de horas de aplicar distintos tratamientos para reanimarla, comenzó a respirar intermitentemente, su corazón volvió a latir y al otro día se recuperó del coma. Este es uno de tantos casos de reanimación de personas con hipotermia debido a inmersión en aguas heladas, confinamientos en la nieve (por ejemplo, tras un alud), o personas que viven en situación de calle y que no resisten una noche de frío intenso[235].

Tanto la niña sueca, como otros que recuperaron sus signos vitales tras varias horas, siguen siendo los mismos de siempre. La pérdida de la consciencia por el congelamiento de su cuerpo no hace que revivan con otra identidad. Cabe aclarar que en este caso la temperatura del cuerpo no baja de los 10 grados Celsius, lo cual dista mucho de los 196 grados bajo cero a los que se crionizan a las personas en Alcor. Más allá de los problemas técnicos relacionados con la vitrificación,

aceptemos por el momento que el hecho de conservar un cuerpo a 10 grados o a 196 grados bajo cero no debería cambiar nada en cuanto a la identidad de la persona.

Pero ¿qué diríamos de un alpinista, como George Mallory, quien desapareció escalando el Everest en 1924 y su cuerpo fue encontrado 75 años después[236]? Por más que lo descongelemos y le hagamos masajes cardiorrespiratorios, no podríamos resucitarlo (parece utópico, aunque esta es una tarea relativamente fácil comparada con resucitar un cuerpo crionizado a 196 grados bajo cero). Pero supongamos que en un par de siglos, si la reanimación de cuerpos congelados fuera posible, se encuentra el cuerpo (hasta ahora perdido) de Andrew Irvine, quien desapareciera con Mallory en la expedición de 1924. Aplicando los últimos avances de la ciencia, imaginemos que los científicos del futuro recuperan el ritmo cardíaco, la respiración y los reflejos de Irvine, y que tras un tiempo logran resucitarlo. ¿Seguiría siendo él? ¿Sentiría que acaba de despertar de un coma prolongado? Y si argumentáramos que no es el mismo, ¿cuál es la diferencia con la chica sueca? ¿Por qué seguimos siendo los mismos tras un día y no tras varios años congelados?

Andrew Irvine (izquierda) y George Mallory (derecha), dos alpinistas desaparecidos en 1924 tratando de llegar por primera vez a la cima del Everest.

Quizás una diferencia fundamental entre los casos de hipotermia y el de los de escaladores congelados es que los pacientes con hipotermia, por más que no mostraran signos vitales (respiración, actividad cardíaca y distintos reflejos), muy probablemente hayan tenido actividad cerebral antes de ser revividos[237]. En toda nuestra vida el cerebro nunca para, las neuronas no dejan de estar activas, ni siquiera en el sueño más profundo. Este no es el caso de los alpinistas congelados en el Himalaya o de aquellos que esperan revivir en los tanques de Alcor. En su caso, la continuidad del cerebro ha sido interrumpida y no es para nada obvio si mantendrían o no su identidad al ser eventualmente resucitados.

Sigamos discutiendo este problema, pero analizando otra variante de perpetuarnos que nos ofrece la ciencia ficción. Para ello, retomemos por un momento la paradoja de Teseo (ver capítulo 7), usando una variación muy interesante propuesta por Thomas Hobbes

(1588-1679). Imaginemos que a medida que se restaura el barco de Teseo, alguien se dedica a juntar los tablones desechados y vuelve a reconstruir la nave en un puerto lejano con las piezas originales. ¿Cuál es entonces el barco de Teseo? Si nos ponemos en el lugar de los habitantes de la antigua Grecia, muy probablemente consideraríamos que el barco de Teseo sigue siendo el que zarpa todos los años a la isla de Delos. Para ellos, el hecho que alguien haya juntado los tablones originales y reconstruido el barco en otro puerto no cambiaría nada; no sería más que una curiosidad.

Esta es la base del *funcionalismo*: lo que importa es la función de las cosas, más allá de aquello que las constituye. Aplicado al cerebro, el funcionalismo afirma que los estados mentales están determinados por su función y no por el sustrato material que los genera.

Imaginemos entonces que algún día eventualmente podamos no conservar o crionizar, sino reproducir nuestro cerebro[238]. O sea, en vez de mantener inerte por siglos nuestro cerebro tal cual es, consideremos la posibilidad de copiar cada una de nuestras neuronas y las conexiones entre ellas. Esto es, por cierto, una tarea monumental, ya que tenemos alrededor de cien mil millones de neuronas y cada neurona conecta con otras diez mil[239]. La cantidad de posibles conexiones va mucho más allá de lo que podríamos tratar en nuestros días, pero imaginemos que en un futuro cercano la ciencia pudiera lograrlo. Para ser más concretos, imaginemos que reproducimos toda la arquitectura del cerebro en una supercomputadora, es decir, reemplazamos las neuronas por transistores en diminutos chips y los conectamos exactamente igual que en la configuración original. Siguiendo los principios del funcionalismo, el hecho de que la actividad "cerebral" esté generada por circuitos de carbono (materia orgánica) o de silicio (el material inerte usado para fabricar los chips de las computadoras) no debería cambiar nada y consecuentemente la configuración de esta supercomputadora almacenaría nuestra memoria e identidad. Pero ¿seríamos realmente nosotros?

La idea de hacer una réplica sintética de una persona (y la posibilidad de trascender a partir de ello) ha sido muy tratada en la ciencia ficción[240]. Sin duda, una de las frases más populares entre los

fanáticos del género es “Beam me up, Scotty”, el pedido de ser teletransportado que el capitán Kirk le hace una y otra vez a Montgomery Scott, jefe de ingenieros de la nave Enterprise, en la serie de televisión *Star Trek*. El principio de teletransportación consiste en desintegrar a la persona en el lugar de origen y reintegrarla casi instantáneamente en el lugar de destino[241]. Hasta aquí todo bien. Pero el episodio quinto de la primera temporada (que salió al aire en 1968) plantea la siguiente situación: debido a una falla en el teletransportador, se genera una copia del capitán Kirk con una personalidad completamente distinta. La réplica tiene un comportamiento extraño y no dudamos en reconocer al impostor. Sin embargo, en un episodio treinta años más tarde (“Second Chances”[242]) el problema se vuelve más interesante. En este caso, el teniente Riker se encuentra con una copia de sí mismo, que desde hacía ocho años aguardaba ser rescatado en el planeta Nervala IV. Tal encuentro se debe a una falla del teletransportador, que en una visita anterior a ese planeta llevó a Riker de regreso a la nave, pero sin desintegrar al que permaneció en Nervala. ¿Quién es quién? Ambos Riker tienen argumentos igualmente válidos de ser el original, uno por haber continuado su vida rutinaria en la nave y el otro alegando que lo dejaron abandonado. El nudo argumental de este episodio justamente juega con el hecho de que es imposible decidir entre ambos.

Teletransportación. El cuerpo de una persona es desintegrado en el lugar de origen e instantáneamente reintegrado en el lugar de destino.

Tan interesantes son los argumentos y dilemas que genera la teletransportación, que han atraído la atención de filósofos contemporáneos. Uno de ellos es Derek Parfit, quien justamente hace uso de la teletransportación para plantear un dilema sobre la identidad[243]. Primero considera el siguiente caso: imaginemos que entramos en el teletransportador y, al apretar un botón, la máquina escanea la información de todas las partículas que componen nuestro cuerpo (y nuestro cerebro) y luego lo destruye en el punto de partida, para reconstruirlo en el punto de llegada, usando la información transmitida por el teletransportador[244]. Estamos a punto de hacer nuestro primer "viaje" en el teletransportador. A pesar de sentirnos algo nerviosos, queremos a toda costa conocer Marte; finalmente la curiosidad puede más que nuestros miedos y apretamos el botón. Poco después aparecemos en una colonia en Marte y miramos fascinados el paisaje rojizo de nuestro alrededor, alegrándonos de que todo haya funcionado. Pero ¿funcionó?

Para hacer más evidente el problema que implica la teletransportación, Parfit propone un segundo caso: años después repetimos la experiencia; volvemos al teletransportador, apretamos el

botón, pero esta vez no pasa nada, seguimos en la Tierra, esperando ser teletransportados. Salimos del aparato y le avisamos al técnico a cargo que la máquina no funciona. Sin embargo, el técnico nos dice que la teletransportación sí funcionó y que ya estamos en Marte. Un error en la máquina hizo que se retardara la desintegración en el lugar de origen, pero no tenemos que preocuparnos porque seremos desintegrados en breve. Con creciente desesperación alegamos que seguimos ahí mismo, en la Tierra, que no fuimos teletransportados. Sin embargo, el técnico nos muestra las imágenes de Marte en donde nos vemos recién llegados. La persona en Marte es exactamente igual a nosotros, actúa como nosotros y puede contestar todas las preguntas sobre nuestro pasado, incluso evocando memorias que nunca compartimos con nadie. Este es un caso muy distinto al de Quaid y Hauser en *Total Recall*, porque el "original" en la Tierra y la "réplica" en Marte no solo tienen el mismo cuerpo sino también la misma memoria. Hay una copia exacta de nosotros en Marte, alguien que será igual de bueno con nuestra familia, que tratará de terminar el libro que hace años tratamos de escribir, pero esto no es consuelo, en el fondo seguimos ahí, en la Tierra, y no queremos morir; queremos ser nosotros y no aquel en Marte quien vea crecer a nuestros hijos y sea testigo de nuestro futuro.

Con este ejemplo, Parfit busca demostrar que la teletransportación es una condena de muerte. El original muere en el proceso de desintegración, mientras que la persona reconstruida en el lugar de destino no se daría cuenta que acaba de ser creada; de hecho, no notaría nada raro y creería ser la misma persona que apretó el botón en la Tierra porque comparte todas sus memorias. En conclusión, si la ciencia algún día permitiera copiar exactamente nuestro cuerpo (y nuestro cerebro), eso no implicaría perpetuarnos. Produciríamos una réplica perfecta de nosotros (la cual no se daría cuenta que es una copia), pero nosotros mismos dejaríamos de existir y no seríamos testigos de lo que haga esa copia el resto de sus días. Y así como dudamos de la teletransportación, es natural también dudar de si mantendríamos nuestra identidad a partir de la crionización o de la réplica de nuestro cerebro en una supercomputadora.

Parfit continúa argumentando que la persona que aparece del otro lado del teletransportador no vivió nuestras memorias originales pero cree haberlas vivido, porque su cerebro (y la memoria almacenada en este) es una copia exacta del nuestro y no hay manera de establecer la diferencia. En conclusión, la réplica tiene mis recuerdos, pero no soy Yo. A partir de este razonamiento, concluye que la memoria no puede explicar la identidad. Por otro lado, vimos en el capítulo 7 que tampoco somos la continuidad de nuestro cuerpo. ¿Qué es entonces lo que determina nuestro Yo?

La única manera que encuentra Parfit de salir de esta encrucijada es concluyendo que el Yo no existe como tal, que es una ilusión. Por supuesto existe el cuerpo, el cerebro y todas las memorias allí guardadas —lo ilusorio es la sensación de identidad, de ser una persona. Para ilustrar esto, considera un club cuyos miembros se reúnen regularmente. Luego de unos años esas reuniones dejan de convocarse, y tiempo después algunos miembros deciden volver a juntarse como lo hacían antes. ¿Es este club el mismo de siempre o es un club nuevo? Para Parfit esta es una pregunta vacía, sin sentido; lo importante es que los miembros volvieron a juntarse y a interactuar entre ellos; el hecho de que lo consideremos el mismo club o no es irrelevante.

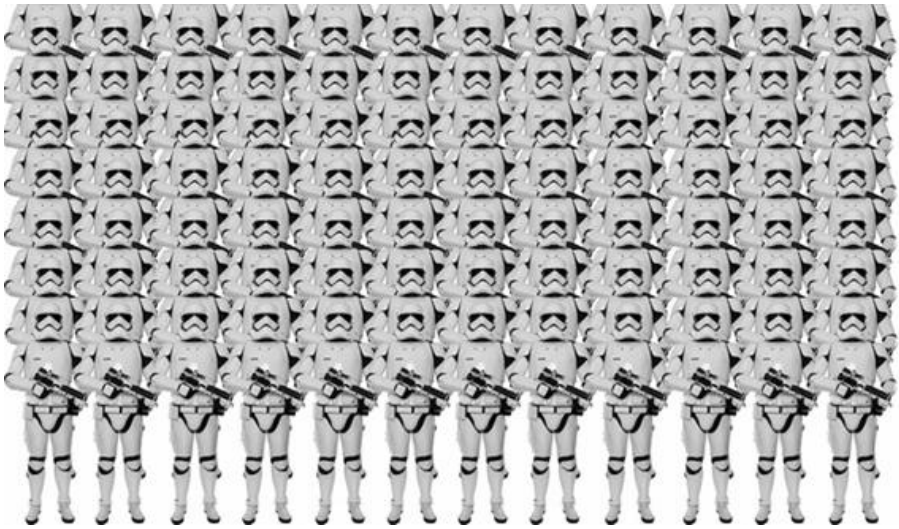
La visión de Parfit no es novedosa. Es el *anatta*, o negación del Yo, que predica el budismo. Cuando un rey alguna vez le preguntó a un monje budista su nombre, este contestó que era conocido como Nagasena, pero que Nagasena era solo un nombre, porque no representa a una persona; según Buda, existen los actos y las consecuencias, pero no la persona que actúa[245]. En la filosofía occidental, la negación del Yo también fue postulada por David Hume (*A Treatise of Human Nature*, 1739, Libro I, parte IV, sección VI), quien alegara:

> Por mi parte, cuando penetro más íntimamente en lo que llamo mi propia persona, tropiezo siempre con alguna percepción particular de calor o frío, luz o sombra, amor u odio, pena o placer. No puedo jamás sorprenderme a mí mismo en algún momento sin percepción alguna, y jamás puedo observar más que percepciones. […]

> [Alguien] Puede, quizá, percibir algo simple y continuo que llame su Yo, aunque yo estoy cierto de que no existe un principio semejante en mí.

Para Hume, el Yo, la identidad, no es más que un manojo de percepciones sucediéndose una tras otra. En esta línea de pensamiento, Dan Dennett afirma que el sentido de identidad es una ilusión creada por nuestra narrativa: es el lenguaje que nos lleva a hablar (y hablarnos a nosotros mismos) y pensar "como si hubiera alguien adentro"[246]. Por otro lado, también vimos que el sentido de pertenencia de nuestro cuerpo es relativo y viene dado por la coherencia de nuestras sensaciones, al creer que una mano de goma puede ser parte de nuestro cuerpo (en la ilusión de la mano de goma) o que otra no lo es, si no se comporta de acuerdo a nuestras predicciones (en el síndrome de la mano ajena). Según esta visión, en el fondo no somos más que sofisticados organismos, pero con tal grado de complejidad que podemos reflexionar sobre nuestras propias conductas y hasta llegar a ser conscientes de nuestra existencia, creando la ilusión del Yo. Y si el Yo no existe, no tiene sentido preguntarse por la muerte porque nunca hubo alguien que vaya a morir.

La negación del Yo es mucho más fácil de aceptar si somos capaces de abstraernos y analizar el problema en tercera persona; si dejamos de pensar qué pasaría con nosotros mismos y vemos todo desde afuera, tomando la perspectiva de un marciano que llega a la Tierra y trata de entender el problema filosófico que nos tiene a maltraer. Volviendo a la ciencia ficción, esta es justamente la postura con la que vemos al ejército de clones de *La guerra de las galaxias*. No nos preguntamos por la identidad de cada uno de ellos; es mucho más sencillo: son un grupo de clones, con el mismo cuerpo y cerebro, viviendo vidas independientes. De la misma manera, la persona que queda en el teletransportador en la Tierra y aquel que llega a Marte —así como el Riker en el Enterprise y aquel que queda en Nervala IV— son copias idénticas que desarrollan sus vidas por separado.



Ejército de clones de *La guerra de las galaxias*. Al no ponernos en la piel de ellos, no nos preguntamos por su vida, por su identidad y su miedo a la muerte.

La noción de que el Yo no existe es una píldora muy difícil de digerir. Es un mazazo a lo que considerábamos la verdad primera de la filosofía —pienso, luego existo— propuesta por Descartes. El argumento sigue siendo válido, pero cambia radicalmente la concepción de qué es lo que existe, pues ya no sería el Yo. Aunque, por otro lado, la sensación de existir, de ser persona, es demasiado fuerte; de hecho, es aquello que creo saber con mayor certeza. No está basado en uno u otro argumento racional; la sensación de que hay un Yo que escribe estas líneas está en mis entrañas. Y así como en un famoso diálogo escrito por John Perry, Gretchen Weirob no se contenta con argumentos generales y descarnadamente se pregunta por la continuidad de su propia vida días antes de morir[247], no podemos evitar analizar el problema en primera persona. Esa es la visión de películas como *Moon*, u *Oblivion*, en donde nos mimetizamos con el personaje principal y su apego a la vida —ya sea una persona que supervisa instalaciones automatizadas para extraer combustibles en la Luna, o un técnico que repara drones controlando generadores de energía en el ya abandonado planeta Tierra— para luego enterarnos que es uno de tantos clones. Estas películas hábilmente nos sacan del planteo en tercera persona de *La guerra de las galaxias* (salvo

a partir del episodio 7, en donde un clon también se plantea por sí mismo) y nos llevan a hacernos las mismas preguntas que se hacen estos clones, las mismas preguntas que venimos haciéndonos desde hace siglos.

Para Parfit, preguntarse por la continuidad de la persona es análogo a preguntarse si un preciado reloj de oro que llevamos a reparar sigue siendo el mismo después de haber sido desguazado y vuelto a ensamblar por el relojero; no hay nada que explicar, la discusión es tan irrelevante como preguntarnos cómo deberíamos considerar al club de un grupo de personas que han vuelto a reunirse. Sin embargo, hay una diferencia fundamental. Tanto el club como el reloj no tienen consciencia de sí mismos. En estos casos no hay otra alternativa que ver el problema desde afuera; no podemos ponernos en el lugar del club o del reloj porque no son seres conscientes. Mas si consideramos una persona, hay algo que sí tiene existencia, aunque solo sea "nominal" (como afirma el budismo). En otras palabras, por más que aceptemos que la noción del Yo es una ilusión —en realidad, preferiría decir una *construcción* del cerebro, así como lo son la mente y el libre albedrío— tiene sentido preguntarnos por aquello que determina el sentimiento que tenemos de nuestra propia existencia, de ser la misma persona a lo largo del tiempo.

La sensación de existir, de ser conscientes de nosotros mismos, es análoga a lo que llamamos *qualia*, la sensación que tenemos de ver el color rojo o de oler un jazmín. Volvemos nuevamente al problema difícil que tratáramos en capítulos anteriores. No percibimos ondas electromagnéticas de una cierta frecuencia, o una determinada composición química de moléculas; *sentimos* el color rojo y el olor del jazmín. Análogamente, no tendemos a percibir un manojo de sensaciones (por más que en la introspección podamos ser conscientes de ellas) sino nuestro Yo. Volviendo a los registros en pacientes epilépticos, en muchos pacientes encontramos neuronas que respondían a distintas fotos de sí mismos; es decir que en un área clave para la memoria y el pensamiento, los pacientes tenían una representación conceptual (y no de detalles o de sensaciones aisladas) de su propia persona.

Lo que a mi entender demuestra Parfit con el ejemplo de la teletransportación es que no necesariamente tiene que existir un único Yo, que pueden convivir infinidad de copias una persona, y que cada una de estas copias tendrá una noción de identidad que irá divergiendo con el tiempo, a partir de las distintas experiencias vividas. Nuevamente, la paradoja se da al analizar el problema en primera persona, tomando el punto de vista del clon de *Moon* o de *Oblivion*. Es entonces que nos preguntamos: ¿qué es lo que hace que Yo este acá y no donde está mi clon si somos exactamente lo mismo? ¿Deberíamos acaso replantearnos el materialismo y aceptar que hay algo más allá de la materia que determina que estemos en uno u otro lugar? La respuesta es que estoy tanto aquí como allá, pero no porque tenga algún tipo de conexión con mi clon, sino porque no hay un único Yo, y analizando el problema desde afuera, en tercera persona, notaríamos que mi clon estaría haciéndose la misma pregunta, considerando que Yo soy una copia. Por otro lado, la ilusión de continuidad de la identidad, de ser una misma persona a lo largo del tiempo, está efectivamente dada por la memoria. Este organismo al que llamo Yo recuerda lo que le aconteció en el pasado y puede planear su futuro en función de ello. A partir de la memoria generamos un relato, nuestra historia. Replicar nuestra memoria no nos haría entonces perpetuarnos, como mostrara Parfit con sus ejemplos de teletransportación, pero es la memoria la que nos genera una visión unificada y coherente de nuestra existencia; es la memoria (por más que inicialmente sea la misma) la que hace sentir persona a cada uno de los soldados del ejército de clones.

Vamos a darle otra vuelta de tuerca al experimento de Parfit. Ya vimos que la falla en el teletransportador hará que el Yo en la Tierra se pregunte por la creación diabólica que apareció en Marte, así como el Yo en Marte, el que vivió la experiencia de ser teletransportado, se preguntará por el "desecho" que quedó en la Tierra. Pero ahora imaginemos que el teletransportador funciona correctamente, aunque sin llevarnos a ningún lado. Simplemente destruye nuestro cuerpo (y nuestro cerebro) y vuelve a construirlo tiempo después, exactamente en el lugar de origen. Imaginemos que desaparecemos y reaparecemos

tras una décima de segundo. ¿Seguiremos siendo los mismos?

Imaginemos, en cambio, que la reconstrucción se hace tras varios años. Curiosamente, esto sería análogo a resucitar a alguien de una crionización; el hecho de que hayamos desmembrado las partículas y que las hayamos reconstruido exactamente como estaban no debería hacer diferencia. Pero antes admitimos que la teletransportación es una condena de muerte y, por lo tanto, tendríamos que aceptar que tras la crionización no seríamos la misma persona. Siguiendo este argumento, ni aunque fuésemos reconstruidos tras una décima de segundo seguiríamos siendo los mismos. Esa mínima interrupción sería nuestra muerte y la aparición de alguien exactamente igual a nosotros, quien a partir de la continuidad de su memoria sentiría que todo salió bien y que la teletransportación no acarrea problema alguno.

¿Qué es entonces lo que importa, qué es lo que mantiene la identidad y evita nuestra muerte? Quizás sea la continuidad de la actividad del cerebro, como especuláramos. Sin embargo, esto es un argumento débil, pues en principio parecería no haber gran diferencia entre interrumpir la actividad del cerebro y alterarla de golpe; y la actividad de cerebro cambia dramáticamente, por ejemplo, si sobrevivimos una descarga de alta tensión, una brutal trompada de nocaut o, sin ir más lejos, cada vez que dormimos. Entonces, la gran pregunta es si no estaremos muriendo constantemente, reviviendo en otro Yo que tiene la ilusión de ser el mismo de siempre a partir de compartir una misma memoria. Aceptar la identidad, incluso como una construcción del cerebro, implica concebir posibilidades asombrosas y lidiar con enigmas para los cuales no necesariamente tenemos respuesta.

Quisiera tocar una última gran pregunta antes de terminar este libro. Reflexionamos hasta ahora sobre aquello que determina la percepción del Yo y la continuidad de nuestra existencia, pero ¿por qué el cerebro genera tal construcción? ¿Existe acaso alguna ventaja evolutiva que nos dé el hecho de tener conciencia de nosotros mismos? A mi entender, la clave es que el sentido de identidad nos lleva a preguntarnos por nuestro futuro y a enfrentar la muerte. El

miedo a la muerte a su vez nos lleva a apegarnos a la vida, incluso tratando de encontrar maneras de desafiar, o al menos prolongar, aquello que sabemos inevitable.

Sobre el final de *Terminator 2*, el modelo T-800 (interpretado por Arnold Schwarzenegger) no duda en destrozar su brazo para librarse de la máquina que lo tiene atrapado, y tampoco vacila en sumergirse en un tanque con metal fundido para destruir el chip en su cabeza[248]. Muy distinta es la reacción de la supercomputadora HAL 9000 en *2001* o del androide Roy Batty (el líder de los replicantes) en *Blade Runner*. En esos casos, la trama busca humanizarlos, que no los veamos solo como máquinas y que los tomemos como seres con un sentido de identidad que se preocupan por su extinción.

En los primeros capítulos nos preguntábamos por aquello que distingue a una máquina de un humano. Más allá de nuestra capacidad única de pensamiento, basados en la abstracción y la extracción de un significado de las cosas, la sensación de ser persona, de tener un instinto de supervivencia —un *conatus*— nos diferencia de las máquinas. En otras palabras, el hecho de que alguien o algo tenga miedo a la muerte, muestra que hay un ser que existe y es consciente de sí mismo.

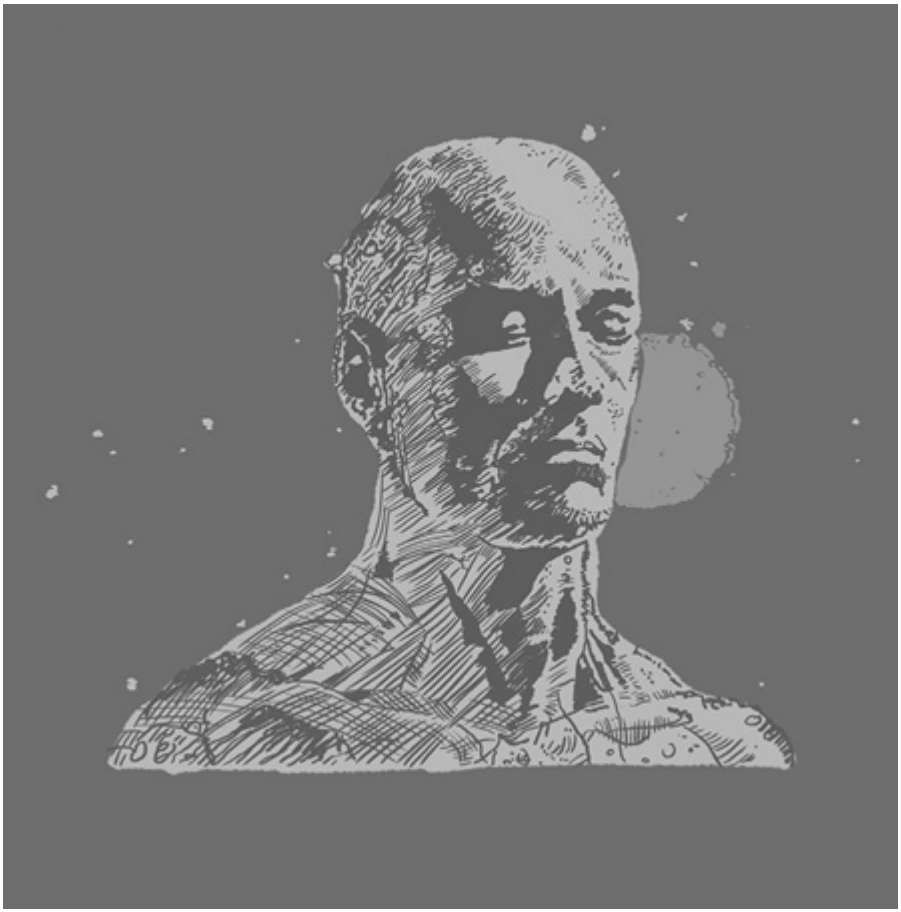
Retomando la discusión de cómo saber si una máquina es consciente (basada en el famoso test de Turing mencionado en el capítulo 2), podemos postular que, tal como HAL 9000, si algún día una computadora demostrara tener miedo de ser permanentemente desconectada (y sin que esto haya sido programado específicamente; es decir, sin que sea la consecuencia de ejecutar una acción establecida en su código), probablemente admitiríamos que es consciente, que puede procesar y entender su contexto y tener *qualia*. Por ejemplo, una computadora programada para reconocer caras o patrones visuales que fuera consciente de sí misma podría temer su final al no querer dejar de sentir y apreciar lo que ve.

Esto también nos diferencia de otros sistemas y organismos. A partir de la interacción de billones de usuarios, Internet tiene una complejidad extraordinaria pero carece de un fin en sí misma. El único motivo de su existencia está dado por las necesidades de los usuarios.

Es como si los distintos órganos y células de mi cuerpo actuaran por sí mismos, sin que hubiera nada que les diera un sentido superior, preservar mi vida. Si en el futuro los usuarios encontraran un medio de interacción que volviera obsoleta a Internet, esta desaparecería sin más, sin la menor disputa o lamento, pues nunca hubo una finalidad superior —una consciencia de identidad y existencia— que llevara a preservar la red en sí misma.

Asimismo, la consciencia, como el miedo a la muerte, es muy limitada en seres inferiores como los insectos. Una avispa no duda un instante en sacrificar su vida al clavar el aguijón, ni se plantea qué será de ella al hacerlo. En nuestro caso esto no es natural, sino que se da en circunstancias extremas, cuando el motivo que nos lleva a arriesgar o sacrificar la vida es tan grande que supera el miedo a la muerte. La consciencia de nuestra propia existencia nos lleva a tratar de preservarnos.

Convengamos entonces que el Yo no es algo mágico y etéreo que nos hace persona, sino que es una construcción del cerebro. Podemos discutir si esta existencia es real o nominal; sin embargo, ilusión o no, la sensación de nuestra existencia es muy fuerte y nos lleva a temer la muerte, a prolongar nuestra existencia y a tratar incluso de inmortalizarnos. Pero más allá de nuestra reacción innata, ¿es esto lo que realmente queremos?



Gilgamesh el inmortal. Imagen adaptada por Candela Insua.

En la historieta de Lucho Olivera que mencionáramos al principio del capítulo, el marciano Utnapishtim cumple su promesa de volver a Gilgamesh inmortal y, tras haber terminado el proceso, le dice que cuando se canse de la vida eterna vaya a buscarlo al planeta rojo. Más de dos milenios antes de Cristo, Gilgamesh le pregunta cómo podría llegar a Marte. Utnapishtim contesta que deberá esperar miles de años, algo insignificante para él, hasta que los desarrollos tecnológicos de su especie lo permitan. Por otro lado, aunque Utnapishtim viaje constantemente por los confines del universo, Gilgamesh tendrá toda la eternidad para encontrarlo.

Gilgamesh, el inmortal, es testigo del desarrollo de la civilización humana y luego debe pasar siglos en total soledad tras su extinción en una guerra nuclear. Cientos de años no representan nada para él, pero

debe esperar el paso de cada año, de cada día y cada hora, hasta reencontrarse con Utnapishtim para liberarse del suplicio de la vida eterna. Por supuesto esto es la ficción de una historieta, pero nos lleva a tomar fuerzas para enfrentar el gran tabú y replantearnos nuestra postura frente a la muerte. El problema es que con la inmortalidad todo pierde sentido: ¿por qué hacer algo hoy, si lo podemos hacer mañana, o en 1000 años?

Sabemos que no somos inmortales, pero tratamos de no darnos cuenta, de ignorar nuestro destino. Y no es casualidad que quienes han visto la muerte de cerca, reponiéndose de un grave accidente o sufriendo una enfermedad con magras expectativas de supervivencia, probablemente hayan cambiado radicalmente su vida tras hacerse una y otra vez las preguntas que tratamos de evitar.

La muerte nos da una urgencia de vivir, le da un sentido a nuestros días. Por más angustia que nos provoque ser conscientes de la fugacidad de nuestro tiempo, la muerte es vida, y en vez de buscar desafiar lo inevitable, quizás sea mejor acostumbrarnos a la idea de que el presente vuela, de que los años pasan para no volver y de que tenemos que disfrutar el tiempo que nos toca vivir, pues tarde o temprano todo acaba en el dulce sueño.

224 Lamentablemente, la editorial Columba, fundada en 1928, cerró definitivamente con la crisis de 2001. Los héroes de nuestra infancia apenas sobreviven en ejemplares desgajados en rincones ocultos de algunas pocas librerías de segunda mano y en los recuerdos de sus epopeyas, alegrando interminables tardes de siesta de verano.

225 Más allá de la posibilidad de trascender la muerte, la ciencia ficción también utiliza esta idea para que tripulantes puedan realizar largos viajes interestelares: si quisiéramos enviar humanos a planetas distantes, podríamos mantenerlos en un estado de hibernación, quizás por décadas o siglos, y hacer que despierten poco antes de llegar a destino.

226 El éxito de la película dio lugar a una remake de Hollywood llamada *Vanilla Sky*, con Tom Cruise y Penélope Cruz (quien también actuó en la versión original).

227 Esta idea nace a partir del libro *The prospect of immortality* publicado por Robert

Ettinger, un profesor de física y matemática de Michigan, en 1962.

228 Para que la crionización funcione (si es que alguna vez realmente funciona) es fundamental que el proceso comience lo antes posible después de la muerte de la persona. Cuanto mayor es el retardo, más grande es el deterioro, particularmente del cerebro, por falta de irrigación sanguínea, como en casos de accidentes cerebrovasculares.

229 Foto cortesía de Alcor Life Extension Foundation.

230 Ver por ejemplo:

- R. Merkle. "The technical feasibility of cryonics". *Medical hypotheses* 39: 6-16; 1992.
- O. Moen. "The case for cryonics". *Journal of Medical Ethics* 41: 677-681; 2015.

231 Este es un proceso muy usado en neurociencia para estudios *in vitro*: se cortan rodajas del cerebro, las cuales son mantenidas en una solución que mantiene vivas las neuronas mientras se estudia durante no más de unas pocas horas su comportamiento a través de registros electrofisiológicos, o registros de imágenes con microscopios especializados.

232 R. Andjus and J. Lovelock. "Reanimation of rats from body temperatures between 0 and 1º C by microwave diathermy". *The Journal of Physiology* 128: 541-546; 1955.

233 Robert McIntyre and Gregory Fahy. "Aldehyde-stabilized cryopreservation". *Cryobiology* 71: 448-458; 2015.

234 Las personas que yacen en los bancos de crionización suelen ser millonarios debido al alto costo del procedimiento, que no baja de 100.000 dólares para preservar solo la cabeza, y más del doble para preservar todo el cuerpo. Curiosamente, no está claro si estas personas retendrían sus fortunas en caso de ser potencialmente resucitadas, ya que legalmente los muertos no tienen derechos ni pertenencias.

235 El caso de la niña sueca esta descripto en: Birgitta Romlin, Helena Winberg, Magnus Janson, Boris Nilsson, Kerstin Björk, Anders Jeppsson, Gunilla Drake and Andreas Claesson. "Excellent Outcome With Extracorporeal Membrane Oxygenation After Accidental Profound Hypothermia (13.8°C) and Drowning". *Clinical Care Medicine* 43: e521-e525; 2015.

Para la reanimación en estos casos es fundamental que la hipotermia se dé antes que la hipoxia, o sea, que la persona y sus órganos se congelen y entren en un estado de hibernación antes de que la falta de oxígeno comience a provocar la muerte celular. Este es, por ejemplo, el caso de niños que quedaron sumergidos en aguas heladas y recién se ahogaron luego de que la temperatura de su cuerpo bajara dramáticamente. Paradójicamente, las chances de resucitar a una persona poco después de ahogada son mayores si el ahogamiento se produce en aguas heladas que en aguas más cálidas. Para una revisión de estos casos ver: Peter Paal, Les Gordon, Giacomo Strapazzon, Monika Brodmann Maeder, Gabriel Putzer, Beat Walpoth,

Michael Wanscher, Doug Brown, Michael Holzer, Gregor Broessner and Hermann Brugger. “Accidental hypothermia–an update”. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 24: 111; 2016.

236 Aún es motivo de controversia si George Mallory y Andrew Irvine, su compañero de escalada, fueron los primeros en hacer cima en el Everest antes de morir.

237 Esto es algo que en principio no se reporta porque en esa situación crítica lo último que se piensa es en hacer un electroencefalograma.

238 Para simplificar vamos a limitarnos a copiar el cerebro, sin tratar de reproducir también el cuerpo. El cerebro estaría aislado del exterior, pero en principio mantendría nuestras memorias —nuestro Yo— y sería una manera de preservarnos.

239 Considerar solo neuronas que conectan entre sí es también una gran simplificación. Por un lado, hay que separar conectividad anatómica (la que vemos en el microscopio) de conectividad funcional (las conexiones que están activas). Por otro, las neuronas reciben las conexiones de otras en sus dendritas y la localización exacta de tales conexiones dentro del árbol dendrítico influencia sus respuestas. Para más detalles, ver, por ejemplo: Tiago Branco, Beverley Clark and Michael Häusser. “Dendritic Discrimination of Temporal Input Sequences in Cortical Neurons”. *Science* 329: 1671-1675; 2010.

240 Ya aparece en *Metrópolis*, la primera película del género, en donde los dueños de una gran corporación reemplazan a la influyente María para cooptar a los trabajadores.

241 Curiosamente, la idea de la teletransportación en *Star Trek* —una de los recursos más célebres en la ciencia ficción— nace a partir de una irrisoria limitación práctica. Por el diseño de la nave Enterprise, se hacía muy complejo y costoso la filmación de escenas mostrando a la nave aterrizando en distintos planetas. La solución fue entonces teletransportar a los tripulantes y que la nave quedara siempre en el espacio.

Para una interesante discusión sobre los principios físicos involucrados en la teletransportación, ver el capítulo 5 del libro *The Physics of Star Trek,* del físico norteamericano Lawrence Krauss.

242 El episodio 24 de la sexta temporada de *Star Trek: The Next Generation*.

243 Derek Parfit; *Reasons and persons*; Clarendon Press; Oxford; capítulo 10; 1984.

244 Vale aclarar que a fin de evitarnos los problemas que acarrea transportar partículas a grandes velocidades, lo único que se transmite es la información de cómo estas se hallaban ensambladas, replicando nuestro cuerpo a partir de partículas disponibles en el lugar de destino. Funcionalismo puro.

245 Citado de D. Parfit; *Reasons and Persons*; Oxford University Press; p. 502; 1984.

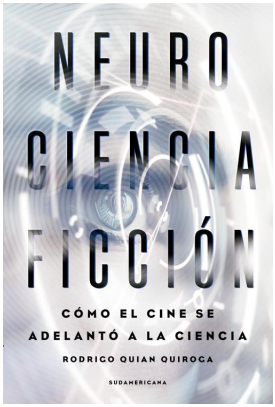
246 Dan Dennett; op. cit.; p. 418.

247 John Perry; *A dialogue on personal identity and immortality*; Hackett Publishing; Indianapolis; 1978.

Gretchen Weirob, una filósofa que yace en su lecho poco antes de morir, discute con

un clérigo y un antiguo estudiante suyo sobre la posibilidad concreta de una vida tras la muerte. A través de este diálogo, Perry revisa críticamente distintas posturas filosóficas y religiosas sobre la identidad personal y su continuidad.

248 Curiosamente en el guion original de la película, Sarah Connor le pregunta a Terminator si tiene miedo justo antes de morir, a lo cual este contesta que sí. Sin embargo, este diálogo no aparece en la película, corrigiendo a último momento lo que hubiera sido una gran incoherencia, ya que al presentar a Terminator como una máquina insensible, no tendría sentido que temiera su muerte.



El cine imagina más allá de lo posible, suele anticiparse a su época y propone avances fascinantes y disruptivos. En *Inception*, un grupo de conspiradores implanta falsos recuerdos; en *Hasta el fin del mundo*, un científico alocado llega a leer los sueños; en *2001: Odisea del espacio*, una supercomputadora siente y piensa como una persona.

Pero así como el cine de ciencia ficción se apoya en los últimos avances de la ciencia, la ciencia también se nutre de la prolífica imaginación de los cineastas y de pronto vemos que es una realidad aquello que alguna vez fue fantasía: implantar memorias manipulando a voluntad grupos de neuronas a través de la *optogenética*, leer la mente durante el sueño a partir de avanzados algoritmos de *decodificación*, o lograr que las computadoras superen el pensamiento humano en infinidad de tareas mediante *redes neuronales profundas*, han sido logros científicos concretos de los últimos años.

El físico y neurocientífico Rodrigo Quian Quiroga, director del Centro de Neurociencia de Sistemas en la Universidad de Leicester, en Inglaterra, propone un cruce entre el arte, la ciencia y la filosofía y se pregunta hasta dónde ha llegado el cine con sus fabulosas especulaciones y cuántas de estas se están materializando en los laboratorios del mundo. Para eso, analiza cómo la ciencia está logrando lo que hace décadas parecía imposible, y cómo estos avances nos llevan a replantearnos las grandes preguntas filosóficas que el hombre viene haciéndose desde siempre.

RODRIGO QUIAN QUIROGA

Estudió física en la Universidad de Buenos Aires, trabajó en los departamentos de

neurofisiología y epilepsia de FLENI, hizo un

doctorado en matemática aplicada en la Universidad de Luebeck, en Alemania, tuvo posiciones post doctorales en Julich, Alemania, y en el Instituto de Tecnología de California (Caltech), y breves estancias en el centro RIKEN de Japón y la Universidad de Nijmegen en los

Países Bajos. Actualmente es profesor y director del Centro de Neurociencia de Sistemas en la Universidad de Leicester, en Inglaterra, e investigador visitante en el centro de investigación Barcelonaßeta, en Barcelona.

Ha tenido posiciones de profesor visitante en Caltech, la Universidad de California Los Angeles (UCLA), la Universidad de Magdeburgo en Alemania, la Escuela

Internacional de Estudios Avanzados (SISSA) en Trieste, Italia, y la Universidad de Buenos Aires.

Sus trabajos sobre percepción y memoria en el cerebro

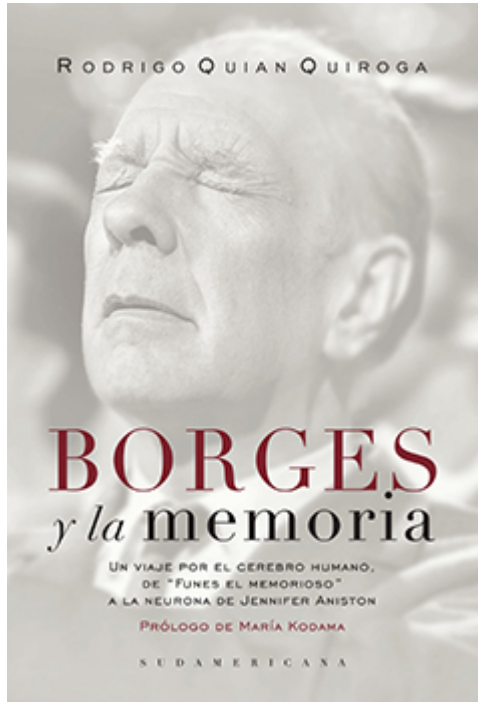
humano han sido publicados en cuatro libros (*Borges y la memoria*, *¿Qué es la memoria?*, *Imaging brain function with EEG* y *Principles of*

neural coding) y en más de cien artículos en las más prestigiosas revistas científicas, los cuales han sido reseñados en *The New York Times*, *The Washington Post*, *The Wall Street Journal*, *Scientific American*, *The Scientist*, *Nature*, *Daily Mail* y *The Independent*, entre otros medios.

Su descubrimiento de neuronas en el cerebro humano

que representan conceptos (las llamadas “Neuronas de Jennifer Aniston”) fue calificado por la revista *Discover* como una de las “Top 100 Scientific Stories of 2005”.

Entre otras distinciones obtuvo el premio al investigador joven por la Sociedad Americana de Epilepsia, el premio al mérito en la investigación científica por la Sociedad Real de Inglaterra, y fue seleccionado por el Consejo de Investigación de Ingeniería y Física y la Real Academia de Ingeniería como uno de los diez científicos líderes en ciencia e ingeniería del Reino Unido.



Otro título del autor en megustaleer.com.ar

Quian Quiroga, Rodrigo
NeuroCienciaFicción / Rodrigo Quian Quiroga. - 1ª ed. - Ciudad Autónoma de Buenos Aires : Sudamericana, 2018.
(Ensayo)
Libro digital, EPUB

Archivo Digital: descarga y online ISBN 978-950-07-6217-5
1. Neurociencias. I. Título.
CDD 612.8

Diseño de cubierta: Penguin Random House Grupo Editorial Ilustraciones de portadillas: Candela Insua Edición en formato digital: noviembre de 2018

© 2018, Penguin Random House Grupo Editorial, S. A.
Humberto I 555, Buenos Aires www.megustaleer.com.ar
Penguin Random House Grupo Editorial apoya la protección del *copyright*.
El *copyright* estimula la creatividad, defiende la diversidad en el ámbito de las ideas y el conocimiento, promueve la libre expresión y favorece una cultura viva. Gracias por comprar una edición autorizada de este libro y por respetar las leyes del *copyright* al no reproducir, escanear ni distribuir ninguna parte de esta obra por ningún medio sin permiso. Al hacerlo está respaldando a los autores y permitiendo que PRHGE continúe publicando libros para todos los lectores.

ISBN 978-950-07-6217-5

Conversión a formato digital: Libresque

Índice

- Neurocienciaficción
- Dedicatoria
- Introducción
- I. 2001: Odisea del espacio
- II. Blade Runner
- III. El planeta de los simios
- IV. Matrix
- V. Hasta el fin del mundo
- VI. Minority Report
- VII. Robocop
- VIII. Inception
- IX. Total Recall
- X. Abre los ojos
- Sobre este libro
- Sobre el autor
- Otro título del autor
- Créditos